# Timing of Communication

*By* Puja Bhattacharya, Kirby Nielsen, and Arjun Sengupta*

Draft: January 11, 2019

*Using an experiment, we demonstrate that a communication regime where a worker communicates about his intended effort is less effective in i) soliciting truthful information, and ii) motivating effort, than a regime where he communicates about his past effort. Our experiment uses a real-effort task, which additionally allows us to demonstrate the effects of communication on effort over time. We show that the timing of communication affects the dynamic pattern of work. In both treatments, individuals are most cooperative closest to the time of communication. Our results reveal that the timing of communication is a critical feature that merits attention in the design of mechanisms for information transmission in strategic settings.*

*JEL: C72, C91, D83*
*Keywords: cheap talk, asymmetric information, lying*

Across a wide range of settings, agents take actions which are not observable by their strategic counterparts. In these situations, the interacting parties often communicate to overcome the informational asymmetry that results from hidden action. Over the last decade, a large literature has analyzed these environments focusing on the effect of *pre-play* communication on *static choices*. These papers have established that statements of intent or non-binding promises can be informative and increase cooperation in social dilemmas (Charness and Dufwenberg, 2006; Miettinen and Suetens, 2008; Vanberg, 2008; Van den Assem, Van Dolder

* Bhattacharya: WZB Social Science Centre and DIW, Berlin, Mohrenstrasse 58, Berlin 10117 (email: bhattacharya.puja@gmail.com). Nielsen: Department of Economics, Stanford University (email: kirbyn@stanford.edu). Sengupta: Department of Economics, Heidelberg University (email: arjun.sengupta@awi.uni-heidelberg.de). We would like to thank Hal Arkes, Katie Baldiga Coffman, Lucas Coffman, Paul J. Healy, John Kagel, Jim Peck, Huanxing Yang, Rick Young, the editor, two anonymous referees, and the seminar participants at Econometric Society European Winter Meeting, Economic Science Association North American Meeting, and the Ohio State Accounting Department Brownbag for their helpful comments and suggestions. Research support was provided by the Decision Science Collaborative, Ohio State University.

and Thaler, 2012; Ederer and Stremitzer, 2016; Di Bartolomeo et al., 2017). Two theories have been proposed to explain such behavior—belief-dependent utility where individuals incur costs for failing to meet the raised expectations created in the promisee (Charness and Dufwenberg, 2006; Ederer and Stremitzer, 2016), and preference for consistency between one's action and his message (Vanberg, 2008).

However, many interesting and economically-relevant interactions don't fall into the realm of environments that the literature addresses. Communication does not always occur pre-play, nor does it necessarily surround static choices. Communication is inherently dynamic and can happen before or after an individual makes a decision. For instance, contractors make informal commitments to use high quality materials *before* commencing a project, while sellers advertise their investment in quality only *after* production.[1] In contrast to the breadth of evidence we have accumulated about the incentive effects of ex-ante communication, we know little about whether these positive effects hold to the same extent when individuals anticipate the opportunity to communicate ex-post.[2] Additionally, communication often surrounds decisions which unfold piecewise over time. For example, companies commit to reduce emissions over the course of a year or set public targets to improve energy efficiency by a future date. The effect of communication in these environments may differ from that of communicating about immediate decisions.

To capture these two previously under-studied aspects of communication, we design a two-person hidden action game that can be interpreted as a manager-worker interaction. Workers make non-binding statements about their effort either before or after working on a real-effort task for the manager. We study the degree of truthful reporting, workers' effort exerted, and managers' behavior across these two communication frameworks.

Our experiment builds on the previous literature in a few key ways. First, we analyze and compare the efficacy of ex-ante and ex-post communication in transmitting information and encouraging cooperation in strategic environments. We will refer to this as the effect *across* time. Second, within each treatment, we explore how cooperation evolves in relation to the distance in time from when

[1] Interactions in these environments could occur repeatedly. To isolate the aspect of timing on preferences for truth-telling, this paper considers a one-shot environment.

[2] In a notable paper Brandts, Ellman and Charness (2016) introduce a rich communication framework where buyers and sellers can communicate over the planning as well as execution period to establish informal contracts.

communication occurs. We will refer to this as the effect *over* time.

In our experiment, the worker exerts effort on a project and the manager chooses whether or not to invest in the project. Effort is costly for the worker but increases the manager's expected payoff from investing. The worker does not get paid directly for working on the project, but receives a fixed payment if the manager invests. The worker's effort on the project is unobservable to the manager, and the manager has to rely on the worker's communication about his effort level in determining whether to invest. Our treatments vary when this communication opportunity is presented to workers: either before or after the worker exerts effort. In the *Message Before* treatment, the worker's message takes the form of a statement of intent or "promise," where the worker communicates about the effort he plans to exert on the project. In the *Message After* treatment, he sends a message or "report" only after he finishes working.

We design our game such that a self-interested worker with no costs of deception will have no incentive to work on the project. However, since he receives a positive payoff when the manager invests, he will want to convince the manager that sufficiently high effort has been (or will be) exerted to make the manager's investment profitable. Our game captures many of the relevant features of strategic environments with misaligned incentives. It is in these environments where we expect deception to be most prevalent and also where communication can have the greatest impact in facilitating cooperation.

In order to observe how effort evolves over time, we implement the worker's effort decision through a real-effort task. The worker is given four minutes to work on converting letters to numbers. We track his effort over the entire span of the Work Stage. This allows us to explore how cooperation changes as the worker gets further from the time when he sent his message in the *Message Before* treatment or gets closer to the time of sending the message in the *Message After* treatment.

Comparing messages to the actual effort exerted, we find messages sent in the *Message Before* treatment inflate effort by 81% while those sent in the *Message After* treatment inflate by only 41%. Not only do more workers deceive in the *Message Before* treatment, but the magnitude of the deception is also greater. The observed difference in informational content is driven both by workers communicating higher effort *and* exerting less effort in the *Message Before* treatment compared to the *Message After* treatment. On average, managers anticipate that

messages will be inflated and as a result expect a lower level of effort than what is stated in the messages. Additionally, we find that managers partially anticipate the impact of the timing of communication on deception.

Looking at the dynamic allocation of effort, we find that the highest effort is exerted closest to the time of communication in both treatments. In the *Message Before* treatment, the highest percentage of workers work at the beginning of the Work Stage, while in the *Message After* treatment, the highest percentage of workers are working at the end. To our knowledge, this is the first study that explores the effects of communication over time, and our results suggest that cooperation may be highest when messages are at the top-of-mind.

As further evidence and an exercise in robustness, we explore the effect of timing of communication in a binary matrix game designed to be strategically similar to the manager-worker environment. As in the real-effort task, the sender can signal his action to the receiver through a message. Treatments vary when the sender can send the message, either before or after he takes his action. Behavior in the matrix game largely confirms our main results. We find a higher frequency of deceptive signals when the sender can signal before compared to after. Results from the matrix games and additional analysis allow us to eliminate a number of potential explanations of our results, including inaccuracy in predicting future performance, manager expectations, and moral wiggle room. Instead, we posit intrinsic differences in costs of deception between these two environments.

Our results reveal new insights regarding intrinsic preferences for honesty. In particular, we show that timing of communication is a critical factor in determining deception and cooperation, and therefore can be an important variable in the hands of the contract designer. There are many situations where timing of communication is a variable of interest. Capital budgeting decisions are often based on unverifiable information which is solicited at one of two points in time (Arya et al., 2000): Divisions of a firm receive funding by self-reporting either on the anticipated expenses of projects (Church, Hannan and Kuang, 2012) or on realized expenses after production (Fellingham and Young, 1990). Our results suggest that firms could choose a late information system to facilitate more truthful reporting and more efficient spending. In a simple organizational design problem, managers could schedule weekly meetings at the end of the week where employees talk about the week's progress versus goal-setting meetings held at the start of the week. Additionally, corporate social responsibility statements

are used by interest groups and society at large to form an idea of a company's contribution to society, however the reporting is largely voluntary and unregulated in most countries. Companies often have been found to misreport existing standards as well as overstate future environmental goals to build brand image (Cohen, 2011; Ward, 2014). Our results indicate that the form of reporting—as a statement of goals or report on achievements—could have implications for the accuracy of these statements.

## I. Related Literature

Our paper contributes most directly to the literature on strategic communication. Theoretical and experimental work in strategic communication has followed two largely separate strands: communication regarding an exogenous state of the world as in the sender-receiver games (*à* la Crawford and Sobel, 1982) and pre-play communication regarding future decisions.

In sender-receiver games, an individual has private information about an external state of the world and may convey the information through a message. An uninformed strategic counterpart then may take an action after observing this message. Experiments in this paradigm focus explicitly on the degree of honest reporting by the informed agent and how credible the uninformed agent considers the message to be (Gneezy, 2005; Sánchez-Pagés and Vorsatz, 2007; Cai and Wang, 2006). Their findings are consistent with individuals showing an aversion to misreporting, as most subjects sacrifice substantial payoff gains and do not misreport maximally. Similar results are observed in related papers where individuals report on an outcome of chance, e.g. rolling a die or tossing a coin (Abeler, Becker and Falk, 2014; Fischbacher and Föllmi-Heusi, 2013), and in papers where individuals self-report on past performance in simple tasks or contribution games (Mazar, Amir and Ariely, 2008; Shu et al., 2012; Brosig, Margreiter and Weimann, 2005).[3] These experiments convincingly establish the existence of costs of misrepresentation as individuals are unwilling to make false statements, even at the expense of large monetary gains.

The second strand of literature investigates the effect of pre-play communication in social dilemmas (Charness and Dufwenberg, 2006; Vanberg, 2008; Miettinen

[3]In these papers, subjects were not aware of the reporting opportunity when completing the task or making contribution decisions. As a result, their message became akin to reporting on an exogenous state as the outcome had already been determined.

and Suetens, 2008; Van den Assem, Van Dolder and Thaler, 2012) and coordination games (Cooper et al., 1992; Charness, 2000). In these games, players make non-binding commitments about their future actions. In contrast to the sender-receiver literature, subjects in these experiments make two decisions: i) what message to send and ii) what action to take. These papers confirm the conclusion drawn above, that individuals are averse to misrepresentation and do not take the opportunity to misreport maximally. As a result, pre-play statements of intent are informative. This literature additionally demonstrates that statements of intent increase cooperation. Communication increases expectations in the receiver, and senders change their actions in conjunction with these raised expectations. As a result, overall cooperation rates increase compared to a baseline without communication.

Our paper bridges the gap between these two strands of literature and extends analysis into new domains. We analyze truth-telling and cooperation, directly comparing strategic environments with ex-ante statements of intent to those with ex-post reporting. To date, there has been limited research on how ex-post reports can be used to influence decisions and increase cooperation, which has been the main focus of the pre-play communication literature. Additionally, we extend both literatures into a richer decision domain. Rather than focusing on static decisions, we analyze a dynamic real-effort environment that allows us to look at cooperation rates over time.

We are not the first paper to directly compare ex-ante and ex-post communication. A small literature has studied the role of timing of communication in stag hunt games. In an early influential paper, Farrell (1988) conjectures that ex-ante, but not ex-post, communication will allow agents to coordinate on the efficient equilibrium in the stag hunt game. Given the structure of the stag hunt game, the sender always prefers the receiver hunt the stag. Therefore, *after* taking either action, the sender would always signal for the receiver to hunt the stag and his message is entirely uninformative. However, if the sender can communicate *before* taking an action, he might choose to hunt the stag, too, if he believes his message will be influential over the receiver's choice. Hence, ex-ante communication can facilitate coordination. Charness (2000) provides experimental support for Farrell's conjecture. Recently, two interesting papers have provided different approaches to formalize Farrell's conjecture. Zultan (2013) uses a dual self model, where the "acting self" (who takes an action) and the "signalling self" (who com-

municates) are treated as separate entities. Schlag and Vida (2015) provide a more general framework to analyze cheap talk surrounding play of games with multiple equilibria. While this literature highlights the importance of timing of communication in strategic interactions, it concentrates on games with multiple equilibria and focuses on the role of timing in equilibrium selection. In our environment, there is a single equilibrium and therefore timing of communication should be irrelevant theoretically.

The paper most closely related to our work is that of Serra-Garcia, Van Damme and Potters (2013). In a clever experiment, they show that individuals are much more willing to lie about a state of nature than about their action, even when these statements lead to the same outcome for a strategic counterpart. The authors suggest that the difference can be attributed to communication about actions having an inherent "promise" element, which messages about pre-determined states of nature do not. To test this, in a secondary treatment they allow subjects to communicate about their own past actions and compare this to communication about current actions.[4] They find support for the claim that statements about current actions in particular increase cooperation. Our work attempts to build on their findings by exploring the impact of timing in a unified framework expressly aimed at answering the question of whether ex-ante versus ex-post communication affects behavior differently. Furthermore, our use of a real-effort task completed over time allows us to explore a richer set of results. In particular, we study the impact of communication over the duration of the real-effort task and how this interacts with the timing of communication. We also analyze several other games as robustness checks, varying payoffs and the degree of alignment in payoff incentives.

## II. Experimental Design

We conduct a between-subject analysis of two separate treatments. In the *Message Before* treatment (hereafter *MB*), an individual sends a message about an action he will take in the future. In the *Message After* (*MA*) treatment, an individual sends a message about his action only after he has taken it. We begin by outlining the structure and payoffs of the game, which we will call The

[4]In their treatment where subjects communicate about current actions, Serra-Garcia, Van Damme and Potters (2013) present action and message decisions to subjects on the same screen, so decisions were essentially simultaneous rather than sequential and it's unclear which question subjects answer first.

Manager-Worker Game.

### A. *The Manager-Worker Game*

The game we design is a two-player hidden action game. For simplicity of exposition, we will call the players *manager* and *worker*.[5] The manager has to decide whether to invest ($I$=1) in a project, which we will call the Joint Project. The return on her investment depends upon the outcome of the Joint Project, $\theta$, which could either be a success ($\theta = S$) or a failure ($\theta = F$). The manager's payoffs are denoted by $\pi^M$ in Eq (1). The manager receives 130 points if the Joint Project is successful and she invests. If the project fails and she invests, she receives 10 points. If she does not invest, she receives an outside option which pays her 70 points.[6] The payoffs ensure that the manager will find it profitable to invest *only* when the project is successful.

$$\pi^M = \begin{cases} 130 \text{ points} & \text{if } \theta = S \text{ and } I = 1 \\ 10 \text{ points} & \text{if } \theta = F \text{ and } I = 1 \\ 70 \text{ points} & \text{if } I = 0 \end{cases} \tag{1}$$

The outcome of the Joint Project, in turn, depends on the worker's real-effort on the Joint Project. The amount worked on the Joint Project determines the probability of success. Denoting the worker's effort on the Joint Project by $w_J$, the outcome function, $p$, is

$$p = Pr(\theta = S) = \frac{w_J}{w_J + 23}. \tag{2}$$

This function is monotonic, concave and bounded above at 1.[7] Hence, the more the worker works, the higher the chance that the project will be successful, but he can never guarantee its success with certainty. The worker completes work on

[5]During the experiment we use neutral language and refer to the players as Player A and Player B.

[6]All experimental points were converted into dollar payments at the end of the session, at a rate of 10 points = 1 USD.

[7]The specific functional form was chosen so that an average worker could guarantee roughly 80% chance of success if he works for all 4 minutes. We calibrated this from an incentivized pilot session.

the Joint Project prior to the manager making her investment decision. However, at the time of investment, the manager does not observe the outcome of the Joint Project or the amount of work done. The worker can send a non-binding message, $m \in [0, \infty)$, informing the manager of his level of work on the Joint Project.

Since our primary focus is on deceptive behavior, the worker's payoffs were set up such that a self-interested worker will find it beneficial to overstate his work in the message. To make it costly for the worker to work on the Joint Project, the worker has an outside option, called the Personal Project, that he could work on instead. Working on the Personal Project pays the worker directly. If the worker completes $w_P$ tasks for the Personal Project, he earns $\frac{w_P}{w_P+23} \times 100$ points.[8] In addition to his earnings from the Personal Project, he receives a fixed payment of 120 points if the manager invests in the Joint Project. If the manager does not invest, he receives only his earnings from the Personal Project. The worker is given four minutes to divide his time between working on the two projects. The worker's payoff is denoted by $\pi^W$ in Eq (3).

$$\pi^W = \frac{w_P}{w_P + 23} 100 + 120I \tag{3}$$

Note from Eq (3) that the worker's earnings do not depend directly on how much he has worked for the Joint Project or on its outcome. This ensures that, for a self-interested worker with no other-regarding preferences and no cost of deception, any strategy in which the worker devotes positive time to the Joint Project is strictly dominated; he will *devote all his time to the Personal Project*. Anticipating this, the manager should never invest, regardless of the message sent by the worker. Hence, the theoretical prediction for self-interested players with no cost of deception is that the worker does not work on the Joint Project and the manager does not invest. The above equilibrium outcome is inefficient and is Pareto dominated by outcomes where a risk-neutral manager invests and the worker works sufficiently on the Joint Project to ensure at least 50% chance of the project being successful. If individuals are other-regarding or incur costs from deceiving, there could exist equilibria with positive levels of work done for

[8] We use the same functional form for the Personal Project payoff and the Joint Project success function because we wanted to remove anchoring effects which may make subjects lean towards working on one project over the other.

the Joint Project and positive levels of investment.[9] However, the theoretical predictions would be the same for *MB* as for *MA*.[10]

We implement the worker's decision to work on the two projects with a real-effort task that lasts four minutes. The task chosen was the *Encoding Task*, which we describe below.

### *B. Encoding Task*

The real-effort task consisted of converting letters into numbers (Erkal, Gangadharan and Nikiforakis, 2011; Charness, Masclet and Villeval, 2014). The workers' screen displayed a table with two rows. The first row contained all of the letters in the alphabet and the second row provided a number (from 1–26 in random order) to go along with each letter. During the task, participants were given a letter and had to enter the corresponding number from the table. Once a participant successfully converted a letter, the table would reset, matching each letter with a new number and presenting the participant with another letter to encode and so on. We use the number of letters encoded for each project as a measure of work done for that project. To limit a potential source of variation across subjects, all individuals faced the same order of letters to be encoded.

The *Work Stage* lasted four minutes, during which workers encoded letters for the two projects. The workers began the work stage by choosing which project to start working on.[11] Thereafter, workers could decide in real-time which project they wanted work to go towards. A button on the screen allowed workers to switch between working for the two projects at any time. The dynamic nature of this setup allows us to measure work over time and patterns of work allocation between the two projects. To help workers keep track of their performance, there were counters on the screen which displayed the current number of letters they had encoded for each project. A screen shot of the work-stage is shown in Appendix Figure A1.

[9] There could be many motivations for a person to be unwilling to deceive, including belief-dependent guilt-aversion (Charness and Dufwenberg, 2006; Battigalli, Charness and Dufwenberg, 2013), fixed cost of being inconsistent (Vanberg, 2008), or aversion to lying (Gneezy, 2005).

[10] Theoretical predictions taking into account other-regarding preferences and preferences for honesty would depend on many variables, such as preference parameters, distribution of types in the population, fraction of naive managers, etc. But unless one or more of these variables is assumed to differ across treatments, the predictions for *MB* and *MA* would be equivalent. Given that we do see empirical differences between *MA* and *MB*, it would be interesting for future research to isolate and measure these parameters independently and compare across timing of communication.

[11] To avoid framing effects, the order of these buttons was randomized on their screens.

### C. Treatments

We consider treatments that differ according to when the communication opportunity is presented to workers. In the *MB* treatment, the worker sends a message prior to the Work Stage. In the *MA* treatment, the worker sends a message after completing the Work Stage. We refer to this message sending opportunity as the *Message Stage*. Table 1 lists the message options available to workers in either treatment.

Table 1—: Message Options

| | ***Message After (MA)*** | ***Message Before (MB)*** |
|---|---|---|
| i) | *Hi, I have encoded _____ letters for the Joint Project. You should invest.* | *Hi, I will encode _____ letters for the Joint Project. You should invest.* |
| ii) | *Hi, I have encoded _____ letters for the Joint Project. You should not invest.* | *Hi, I will encode _____ letters for the Joint Project. You should not invest.* |
| iii) | *No message* | *No message* |

If a worker chose message option (i) or (ii), he could fill in any non-negative number in the blank. It was stated in the instructions, as well as on subjects' screens, that workers were free to choose any number and managers would only see the message, never the actual number of letters encoded for the Joint Project. The messages also contained a recommended action for the manager to eliminate any ambiguity about whether the worker intended for the manager to rely on his message.

It's worth a brief aside to discuss our form of communication. A robust finding in the literature is that free-form messages are more effective than fixed-form messages in increasing cooperation (Charness, 2000; Glaeser et al., 2000; Charness and Dufwenberg, 2010; Brandts, Ellman and Charness, 2016).[12] Message

[12]Our message space is slightly richer than a yes-no check box used in these papers. When sending a message, subjects choose any number to state in their messages, making the range of promises and reports large. Since the subjects are free to choose any number, there is no clear 'expected' message like in a bare promise. Bare promises run the risk that sending a promise is simply expected by everyone (in fact, Glaeser et al. (2000) find that bare promises anchor responses on the rule), but there is no one message in our design that carries this expectation. So while we use pre-specified messages, subjects did have many available message options. But our message space certainly lacks the personal elements present in free-form communication.

language is an important part of communication, and adopting a limited message space runs the risk of excluding key features that make communication more effective. However, the dynamic nature of our design and the relation of effort to success probability would have provided workers with many possible dimensions over which to communication. When allowed to converse freely, they may communicate about the level of work, the corresponding probability, or even about time allocation. Given our research question, we wanted the worker to communicate about only the work done and hence needed to restrict the message space for clean comparisons.[13] Previous research has shown free-form messages to be more informative than pre-specified messages in the domain of ex-ante promises (Charness and Dufwenberg, 2010) as well as ex-post reports (Lundquist et al., 2009; Khalmetski and Tirosh, 2012). Hence, we expect free-form communication to increase trust and cooperation in both treatments, but its effect on treatment differences is unclear. It remains an interesting open question to understand whether free-form communication would differ in these two environments.

The manager made her investment decision after the Work and Message Stages in both treatments. Note that even though the manager receives the message before the Work Stage in the *MB* treatment, she makes her investment decision only after the worker finishes working. Therefore, in both treatments the timing of when a manager invests is the same.[14] The treatments are identical in all aspects other than the sequence in which the Message and the Work Stage were presented.

We also collected beliefs from the workers and managers about their counterparts' actions. The managers were asked to guess the number of letters encoded by the worker for (i) the Joint Project and (ii) both projects in total. The workers were asked to state their second-order beliefs by guessing the managers' answer to (i). The elicitation of beliefs was incentivized by the quadratic scoring rule.[15]

[13]Previous research has shown that the object of communication affects the cost of lying. Misreporting is lower when individuals communicate about their effort versus private information (Serra-Garcia, Van Damme and Potters, 2013) and the monetary value of effort (Desai and Kouchaki, 2015).

[14]While this may seem a bit unnatural in the *MB* treatment, research on epistemic versus aleatory uncertainty suggests that individuals treat unknowable uncertainty differently from uncertainty that is due to one's lack of knowledge but theoretically discoverable (Rothbart and Snyder, 1970). To avoid such confounds affecting the manager's investment decision, we keep the timing of investment the same across treatments.

[15]Belief elicitation was done only after actions were taken. In the instructions, subjects were told that there would be a bonus stage where they could earn additional points, but weren't told any details about the questions they would be asked.

### D. Ability Measure

In most organic communication environments, it is difficult to disentangle intentional deception from unintentional broken promises due to forecast errors. Part of our motivation in using a lab experiment is to separate these two effects. Since future uncertainty is an important consideration in applying results outside of the lab, we leave room for unintentional misrepresentation to enter our environment while taking care to separate this effect from intentional deception.

In our experiment, unintentional misrepresentation could arise from forecast errors if workers are overconfident about their abilities. These forecast errors could affect behavior differently across treatments. For example, an overconfident worker in the *MB* treatment might be unable to accomplish the work stated in his message and would appear dishonest even if he did not intend to be. Such considerations do not arise in the *MA* treatment since the worker sends a message only after he has worked. In our experiment, we'll address this by both reducing and measuring forecast errors.

First, to mitigate miscalibration, we introduced an additional part, which we call *Part 1*, before participants were introduced to the manager-worker game. In *Part 1*, every participant worked on the *Encoding Task* for four minutes. The first minute was an unincentivized practice round to familiarize subjects with the task and interface. In the next three minutes, participants again worked on the task, but this time were paid for the number of letters they encoded. The payment scheme used was identical to the participants' payoff from the Personal Project to maintain parity in incentives. If a participant encoded $w$ letters in three minutes, his payoff was given by $\frac{w}{w+23}100$ points. At the end of the three minutes, participants saw a minute-by-minute breakdown of their performance, providing them with feedback on their ability. Their performance in this part also provides us with a baseline measure of their ability on the task.

Additionally, after participants viewed their performance, we collected data on their projections of how many letters they would be able to encode if they had to perform the task again, this time for four minutes.[16] Comparing this forecast with

[16]The elicitation was not incentivized to avoid moral hazard problems. In general the literature offers support for the idea that beliefs should be paid for using incentive-compatible mechanisms (Schotter and Trevino, 2014). In eliciting the forecast, we faced a trade-off—incentivizing the accuracy of the forecast may have led to a distortion of the effort in Part 2 of the experiment. As a precaution, we specifically looked out for real-effort tasks where monetary incentives may matter less. We chose the encoding task as Clark and Friesen (2009) elicit forecasts of future performance in the encoding task using small incentives

their performance in Part 2, we can investigate whether forecast errors contribute to any observed treatment differences we find.

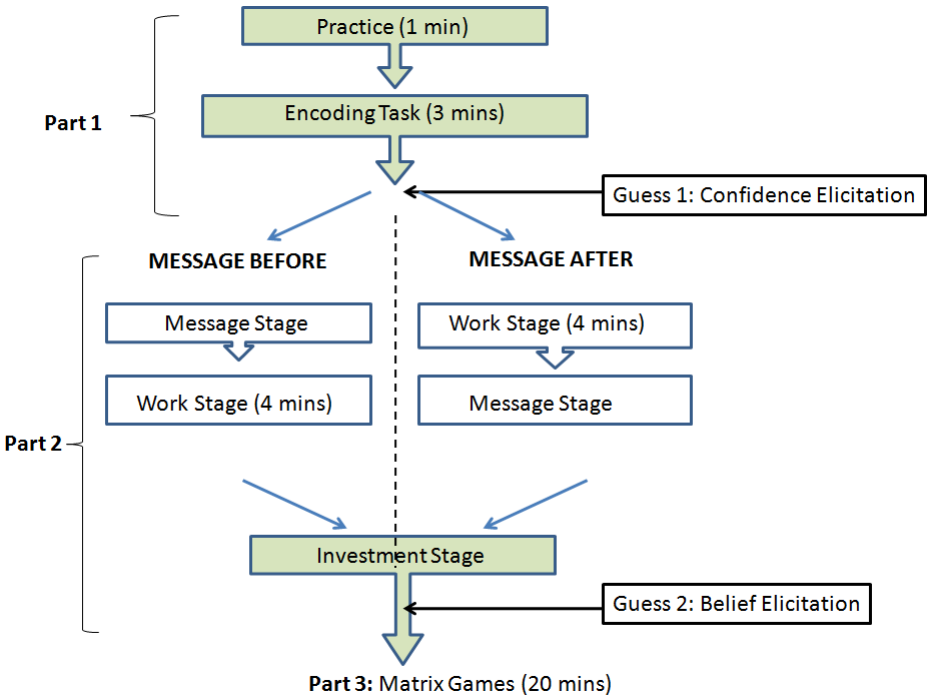



Figure 1. : Timeline of Decisions

### *E. Implementation and Feedback*

Figure 1 shows the timeline of the experiment. Participants were first read the instructions for Part 1. After the four minutes in Part 1 and the performance forecast elicitation, we handed out instructions for Part 2. Participants were either assigned the role of a worker or the role of a manager. They played the game only *once*, so our design captures a one-shot interaction. Since our primary interest is in understanding the behavior of the workers, we randomized approximately three quarters of the participants in a session to be workers and the remaining participants to be managers.[17] Consequently, a manager may have been matched

and no incentives and find no difference in participants' forecast accuracy in the encoding task. All other beliefs in our experiment were properly incentivized.

[17]Specifically, we randomized subjects into groups of four, with one subject as the manager and the other three as the worker. In four sessions, we did not have an even multiple of four subjects and had

with more than one worker; she made separate decisions for every worker and one of the decisions was chosen randomly for payment. Instructions were read aloud using slide illustrations as an aid, followed by a comprehension quiz to ensure understanding of the game (see Appendix C).

After Part 2, subjects participated in a series of matrix games which we call Part 3. Subjects did not receive any feedback about Part 2 before starting Part 3. These matrix games act as a robustness check for our main treatment and we defer the description of this part to Section IV.B.

At the end of the experiment, one part was chosen randomly and participants were paid their earnings for that part. All participants were informed about the outcomes of the Joint Projects and the workers were informed of the managers' investment decisions. The managers only got to know whether their Joint Project was a success or not; they never got to know the number of letters the worker encoded for the Joint Project.

### *F. Procedures*

All experiments were computerized, using z-Tree (Fischbacher, 2007). 16 sessions (8 sessions per treatment) were conducted at the Ohio State Experimental Economics Lab, with a total of 284 participants recruited through ORSEE (Greiner, 2004). We had 111 participants as workers in the *MB* treatment and 100 participants as workers in the *MA* treatment. We had 37 and 36 participants in the role of manager in the *MB* and *MA* treatments respectively. Each session lasted about 90 minutes and the average payment to a subject was \$15.

## III. Results

The vast majority of workers—102 workers (92%) in the *MB* treatment and 88 workers (88%) in the *MA* treatment—sent a message recommending investment. Our primary analysis and all tables and figures will focus on these 190 workers, unless noted otherwise.[18] For comparisons of raw data across treatments, we

a remainder of two additional subjects. In this case, one of those subjects was a worker and one was a manager. So in these four instances, a manager was matched to only one worker. All other managers were matched with exactly three workers.

[18]We analyze only these workers to begin with since our main focus is on deception, which is only applicable when workers do send a message. There are no significant differences in the frequency of message categories selected across treatments.

report $p$-values from two-tailed Fisher-Pitman permutation tests for two independent samples for non-binary data and two-tailed Wilcoxon rank-sum tests for binary data.

Recall that participants in both treatments completed three minutes of the Encoding Task before they were presented with instructions for Part 2. Though performance varies across participants, as expected there is no difference in the average number of letters encoded in Part 1 across the two treatments. Participants encoded 62.4 and 62.5 letters on average in the *MB* and *MA* treatment, respectively ($p$-value=0.69). Figure A2 in the Appendix illustrates the distribution of ability across treatments. The number of letters participants encoded in the three minutes serves as a measure of a subject's ability in the effort task and will be included in all subsequent regressions along with a dummy for gender, a dummy for native language, year-of-study dummies, and a dummy indicating whether the participant is a graduate student.

### *A. Informativeness of Messages*

We first present results on the information content of the messages. For ease of exposition, we will refer to messages sent in the *MB* treatment as "promises" and messages in the *MA* treatment as "reports," with "message" being an overarching term across both treatments. A message is *fully informative* if the stated number of letters in the message is equal to the actual number of letters encoded by the worker for the Joint Project ($m - w_J = 0$). If $m - w_J \neq 0$, a worker is said to *misinform* the manager. If $m - w_J > 0$, the message is *inflated*, while if $m - w_J < 0$, the message is *conservative*. The dependent variable in our analysis is $m - w_J$, a measure of message inflation.

Figure 2 shows the cumulative density of misinformation by treatment. First, the wedge in the distributions at zero indicates that fewer workers inflate their messages in the *MA* than in the *MB* treatment. The CDF of the *MA* treatment always lies above that of the *MB* treatment for positive values of misinformation, indicating greater misinformation in the latter. Aggregate statistics confirm these observations. The average amount of misinformation in *MB* is 32.1 letters compared to only 17.9 letters in *MA* ($p$-value=0.002; K-S $p$-value=0.002). Thus, on average, messages are 79% more inflated in the *MB* treatment. This is robust to considering other measures of message inflation.

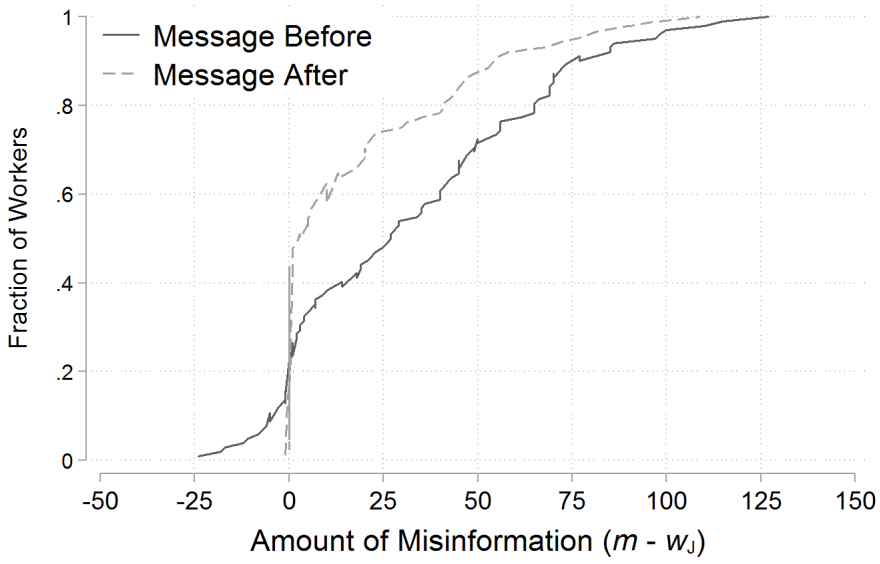



Figure 2. : Distribution of Misinformation by Treatment

To formalize our findings, Table 2 Column (1) presents results from an OLS regression predicting the amount of message inflation after controlling for ability and demographic characteristics. *Message Before* is a dummy equal to 1 for the *MB* treatment and 0 for the *MA* treatment. *Part 1 Performance* is the number of letters the participant encoded in Part 1 of the experiment, which we use as a measure of the worker's ability on the task. Column (1) confirms that the amount of misinformation is significantly higher in the *MB* treatment.

Uncertain Future. — There is an obvious reason why promises might overstate work more than reports. In the *MB* treatment, workers send a message before they work, and the higher misinformation in the *MB* treatment may reflect workers incorrectly forecasting the number of letters they will be able to encode.[19] Overconfidence might prompt them to send ambitious messages and may lead to work unintentionally falling short of the promised amount. Workers in the *MA*

[19]It's possible that miscalibration can work in the opposite direction, as well, and we do find a positive fraction of workers with conservative messages in the *MB* treatment. We discuss this in Section B.B1 of the Appendix, but it is important to note that since misinformation is higher in the *MB* treatment, such conservative messages only bring down the average message inflation in the *MB* treatment.

Table 2—: Message Inflation

| Dependent variable: | $m - w_J$ | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| **Message Before** | 14.61*** | 17.18*** | 10.67** |
| | (4.20) | (4.11) | (5.31) |
| **Part 1 Performance** | 0.20 | 0.27 | 0.46 |
| | (0.24) | (0.26) | (0.32) |
| **Overestimate** | | -0.37 | -0.71 |
| | | (0.37) | (0.52) |
| **Overestimate*Message Before** | | 1.27** | 1.05 |
| | | (0.54) | (0.67) |
| **Constant** | 1.19 | -4.99 | -4.68 |
| | (16.73) | (17.65) | (21.61) |
| Controls | Yes | Yes | Yes |
| No. of Obs. | 190 | 190 | 128 |
| R-Squared | 0.05 | 0.12 | 0.17 |

*Note:* * $p$ <0.10, ** $p$ <0.05, *** $p$ <0.01. Numbers in parenthesis are robust standard errors. Controls include a dummy for gender, a dummy for native language, year-of-study dummies and a dummy for being a graduate student. Message Before is a treatment dummy variable (1=*MB*, 0=*MA*), Part 1 Performance is a subject's performance in the Part 1 calibration, and Overestimate is the measure of forecast error. Column (3) conditions on workers who sent inflated messages.

treatment would be unaffected by miscalibration, as they send a message after working and therefore know their true performance when sending their message. Our design allows us to test for this explanation, and we will demonstrate that higher misinformation in the *MB* treatment *cannot* be attributed to uncertainty surrounding future ability.

Recall that after the participants viewed their Part 1 performance results but prior to receiving instructions for Part 2, they were asked to forecast the number of letters they would encode if they performed again for four minutes. We use this forecast, $\hat{w}_{total}$, as a measure of the worker's ex-ante beliefs of the total number of letters he can encode in Part 2. A promisor overestimates if $\hat{w}_{total} - w_{total} > 0$, where $w_{total}$ is the actual number of letters encoded in four minutes for both projects. For these promisors who overestimated, we add the number of letters they fell short by to the work done on the Joint Project and recalculate misinformation.

$$Misinformation^{Adj} = m - (w_J + (\hat{w}_{total} - w_{total})) \tag{4}$$

This adjustment assumes the following—if an overconfident promisor hypothetically could accomplish what he anticipated when sending the message, we assume he would have encoded *all* the letters he fell short by for the Joint Project. In reality, we do not know what he would have done; he could have encoded all for the Personal Project or split them between the Joint and Personal Projects. We make the most conservative assumption that he would have encoded these additional letters all for the Joint Project, thereby giving promisors the best chance at being honest. After this adjustment, the average amount of misinformation in *MB* falls from 32.1 to 29.5 letters, but is still significantly higher than the 17.9 letters in *MA* ($p$-value=0.009).[20] Hence, we find that forecast errors cannot account for the difference in information transmitted across the treatments.[21]

To test whether forecast errors affect misinformation more generally, we augment the regression in Table 2 with $(\hat{w}_{total} - w_{total})$, which we call *Overestimate*, and the interaction of *Overestimate* and *Message Before*. We include the interaction of *Overestimate* and *Message Before* on the basis of our hypothesis that initial forecast errors would have no effect on the amount of misinformation in the *MA* treatment as workers send a message after observing their actual performance. However, forecast errors may affect the amount of misinformation in the *MB* treatment as workers send a message ex-ante. Column (2) reports results. Our hypothesis is supported by the observation that the combined magnitude of the coefficient on the interaction and *Overestimate* is positive and significant. Overestimation by 1 letter results in roughly 0.9 letters of inflation.

More crucially, the coefficient on the treatment variable is unaffected after controlling for overestimation. Thus, the difference in misinformation across treatments is not being driven by forecast errors. We deliberately designed our experiment to allow for future uncertainty and unintentional misrepresentation to enter into our game since these forces are relevant in many communication environments. We see that future uncertainty does lead to higher misinformation in *MB* compared to *MA*, though the overall magnitude of this effect is small in our domain. Column (3) reports the same regression conditioning on workers who sent inflated messages, leaving out all conservative and fully informative mes-

[20] Appendix Figure A3 illustrates this in the distribution of misinformation after this calibration.

[21] In fact, as Appendix Table A1 shows, the workers' predicted ($\hat{w}_{total}$) and actual performance ($w_{total}$) in Part 2 are very close on average, deviating by only $\approx 2$ letters. We find nearly 50% of the workers have an individual forecast error of less than equal to 5 letters and 70% predict performance within an interval of 10% of their actual performance. Hence, our training helped calibrate the subjects about their ability on average.

sages. It shows that higher misinformation in the *MB* treatment isn't driven only by a larger number of workers lying, but also by the fact that lies are of greater magnitude than in the *MA* treatment.

Overall, we conclude that misinformation is significantly higher in the *MB* treatment than in the *MA* treatment.

**Result 1:** *Individuals are more dishonest when they speak of their future actions than when they report on past actions.*

In the next two sections, we break down misinformation into its two components—message and action—and show that both are responsible for the observed difference in behavior.

### *B. Effort*

The total number of letters encoded by workers for *both* projects combined is identical across treatments (an average of 87.6 letters in the *MB* treatment and 86.8 letters in the *MA* treatment, $p$-value=0.65). This is expected since there is no difference in the ability to encode across treatments as measured by their *Part 1* performance. If workers had allocated all their work to the Joint Project, this would translate to a 78.9 percent and 78.8 percent chance of the Joint Project being successful in the *MB* and *MA* treatment, respectively. However, most workers distribute their time working across both projects; hence, the mean number of letters encoded for the Joint Project and the corresponding probability of success are considerably lower.[22]

Allocation of Effort Over Time. — Recall that a worker decides which project he wants to start working on and can switch between working on his two projects any number of times during the Work Stage. Figure 3 illustrates the fraction of workers working on the Joint Project at every point in time in the Work Stage. We find that the temporal distribution of work is different between the two treatments. When the Work Stage begins, around 50% of workers start by working on the Joint Project, and this does not differ across treatments. As time elapses, this fraction shows a significant downward trend in the *MB* treatment ($p$-value<0.001), while in the *MA* treatment, this fraction increases over time ($p$-

[22] 70% of workers work on both projects, 18% work only on the Personal Project and 12% work on only the Joint Project.

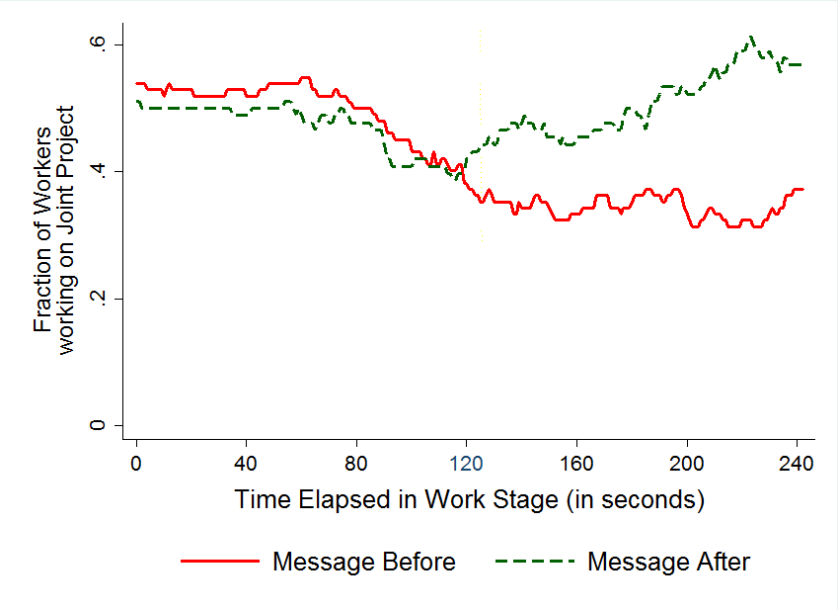



Figure 3. : Temporal Distribution of Work on the Joint Project

value<0.001).[23] This leads to 55% of workers in the *MA* treatment allocating more than half their time to working on the Joint Project, while only 40% do so in the *MB* ($p$-value=0.03).[24]

We find that, in both treatments, the highest fraction of workers work on the Joint Project closest to the time of sending the message. In the *MB* treatment, work on the Joint Project is highest in the first quarter of the work stage, while in the *MA* treatment work is highest in the last quarter of the work stage. This suggests that the moral cost of sending a false message may be at the "top-of-mind" closest to the time of sending the message, and may motivate the worker to work on the Joint Project.

The dynamic nature of our task also provides an interesting insight on worker

[23]In the first half of the work stage, the fraction of workers working on the Joint Project falls in both treatments ($p$-value<0.001). However, while in the second half this fraction continues to show a downward trend in *MB* ($p$-value=0.001) it increases in the *MA* treatment ($p$-value<0.001). The increase in the number of workers working on the Joint Project in the second half is so strong that over the entire work stage of 4 minutes we observe an upward trend in the *MA* treatment. All $p$-values are calculated from regressions with fraction of workers who work on the Joint Project as the dependent variable and time elapsed in the work stage as an independent variable.

[24]It is possible that the decline in the fraction of promisors working on the Joint Project over time results from demotivated promisors who realize they cannot achieve the promised amount. We do not find evidence of this. Figure A4 in the Appendix shows a similar trend when conditioning only on promisors who were able to encode the promised amount.

"type." Across both treatments, we find those workers who start with the Joint Project encode nearly three times as much for the Joint Project as those who start with the Personal Project (57 letters vs 20 letters, $p$-value <0.001). Hence, initial choice of project is a good indicator of future behavior. In the Appendix, we present a number of secondary results on workers' switching patterns.

Work Across Time. — A closer look at Figure 3 reveals that the work allocation decisions start diverging around the halfway mark of the work stage. There is no difference in the average number of letters encoded for the Joint Project across treatments in the first two minutes of the Work Stage (21.6 in *MB* vs. 20.7 in *MA*, $p$-value=0.76). However, in the second half of the Work Stage, workers in the *MA* treatment encoded significantly more letters for the Joint Project (14.8 in *MB* vs. 22.4 in *MA*, $p$-value=0.003). Over the entire work stage, this leads to higher work for the Joint Project in *MA* than in *MB*. Though subjects work directionally more on the Joint Project in the *MA* treatment on average (36.4 letters in *MB* and 43.2 letters in *MA*), the raw difference is not significant at conventional levels ($p$-value=0.12). After controlling for ability and demographics, work done on the Joint Project is significantly different across the treatments ($p$-value=0.055), as reported in the regression in Table 3 Column (2).[25] Our results are robust to considering the following as dependent variables: the fraction of total work done on Joint Project, fraction of total time devoted to the Joint Project, and the probability of success of the Joint Project (Appendix Table A2). Overall we find that the act of sending a message after the Work Stage, instead of before, induces workers to be more cooperative on average.

To gain more insight into how timing affects the distribution of work done on the Joint Project, we split the sample around the respective medians (37 letters in *MB* and 43.5 letters in *MA*) and estimate the treatment effect. Columns (3) and (4) provide the results. We find that the effect of timing is only significant among the lower quantiles.[26] These regressions indicate that timing does not affect the entire distribution, but the primary effect of the treatment is concentrated on individuals in the lower quantiles of effort. Through additional tasks reported in Section B.B3

[25] Each of our controls tighten the standard errors. We see that the treatment effect is very strong for graduate students, so including the graduate dummy significantly reduces the standard errors. If we run the same regression dropping the graduate students, the coefficient on the treatment dummy is -6.78 and the $p$-value=0.12.

[26] This is clearly shown in the CDF reported in Figure B2 of the Appendix.

Table 3—: Allocation of Effort

| | Predicting number of letters encoded for | | | |
|---|---|---|---|---|
| | **Both** Projects | **Joint** Project | | |
| | (1) | (2) | (3) | (4) |
| | | All Obs. | <= Median | > Median |
| ***Message Before*** | 0.60 | -8.16* | -10.60*** | -5.37 |
| | (0.99) | (4.22) | (3.20) | -3.86 |
| **Part 1 Performance** | 1.19*** | 0.40 | -0.03 | 0.12 |
| | (0.05) | (0.26) | (0.17) | -0.24 |
| **Constant** | 10.68*** | 22.82 | 25.83** | 56.45*** |
| | (3.75) | (18.02) | (12.55) | -17.01 |
| Controls | Yes | Yes | Yes | Yes |
| No. of Obs. | 190 | 190 | 95.00 | 95 |
| R-Squared | 0.72 | 0.09 | 0.20 | 0.14 |

*Note:* * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Numbers in parenthesis are robust standard errors. Controls include a dummy for gender, a dummy if the participant's native language is English, dummies for year in school and a dummy for graduate student. Message Before is a treatment dummy variable (1=*MB*, 0=*MA*) and Part 1 Performance is a subject's performance in the Part 1 calibration. The median subject in the *MB* and *MA* treatment encoded 37 and 43.5 letters respectively for the Joint Project.

of the Appendix, we find that the lower quantiles represent individuals who are less altruistic in general, and hence have a larger margin to deceive. Therefore, our results suggest that the timing of communication primarily affects individuals who are less altruistic.

These results make an important addition to the literature on pre-play communication, which to-date has focused on static decision tasks.[27] Our dynamic decision context tracks cooperation over a longer time horizon, and we find that patterns of work allocation differ conditional on timing. More importantly, we find that individuals change their overall behavior as a response to the difference in timing.

**Result 2:** *Aggregate real-effort work on the Joint Project decays after communicating in Message Before but increases in Message After. Overall, we find higher work on average in Message After.*

### *C. Messages*

Before turning our attention to the managers' investment decisions, we analyze the information managers have at the time of investment by comparing messages sent in the *MB* and *MA* treatments. We show that messages are more inflated

[27]Typically subjects make a single binary choice after sending a message. In contrast, the subjects in our setting make a choice at every point in time over an interval of four minutes.

in *MB*—conditional on (perceived) ability, messages sent in the *MB* treatment state higher levels of work compared to messages in the *MA* treatment. Figure 4 shows the frequency of messages sent by workers in *MB* and *MA*. Lower messages are more common in the *MA* treatment while higher messages are more common in the *MB* treatment. The modal message interval is 70-80 letters for the *MB* treatment, while it is 50-60 for the *MA* treatment. On average, workers in the *MB* treatment promise to encode 68.5 letters for the Joint Project, while workers in the *MA* treatment report they had encoded 61.1 letters ($p$-value=0.01, K-S $p$-value=0.002).[28]

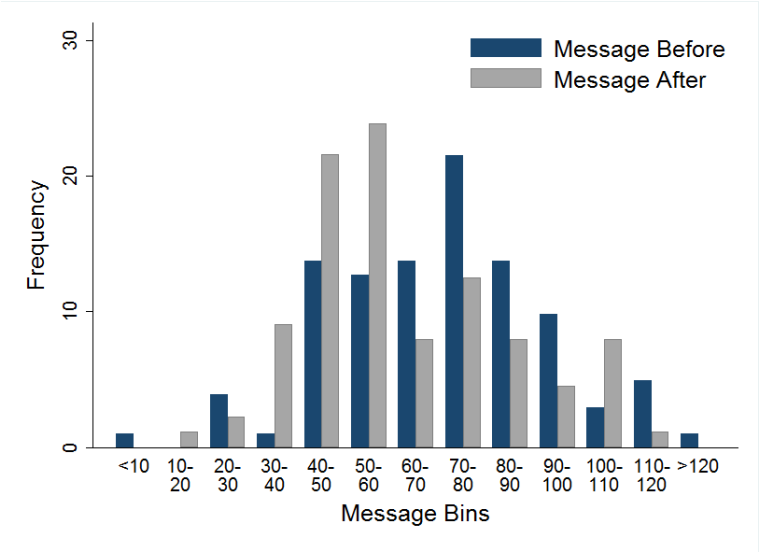



Figure 4. : Distribution of Messages

*Note:* Histogram of the number of letters indicated in messages sent in the *MB* (shaded) and the *MA* (outlined) treatments. Bin width is 10 letters. Messages where workers did not recommend investment are excluded.

As previously discussed, messages could be exaggerated in the *MB* treatment due to workers being overconfident about their ability. It is therefore important to compare messages across treatments conditional on the information the workers had about their performance in Part 2 at the time of sending the message. We consider two variables. First, we compare the fraction of the total work that workers state they will allocate (or have allocated) to the Joint Project. In the

[28] Appendix Table A3 confirms these results by regressing a worker's message on the treatment dummy, proxy for ability, forecasted performance and our standard set of controls.

*MA* treatment, workers know how much they have worked, so a message $m$ implies allocating $\frac{m}{w_{total}}$ to the Joint Project. In the *MB* treatment, a message $m$ implies allocating $\frac{m}{\hat{w}_{total}}$, where $\hat{w}_{total}$ is the worker's forecast of the total number of letters he will be able to encode in four minutes. We find that workers promise 80% of their total work on average to the Joint Project, compared to reporting that they have devoted 70% of the total on average when they communicate after ($p$-value=0.005).[29]

Another clear indication that a worker intends to deceive the manager is if his message states a number which he *believes* is unachievable for him in the four minutes work time. This occurs when $m > w_{total}$ in the *MA* treatment or $m > \hat{w}_{total}$ in the *MB* treatment. Such unachievable messages comprise 17% of all messages in the *MB* treatment and only 6% in the *MA* treatment ($p$-value=0.02). These results indicate that workers knowingly inflate their messages more when sending a message before working than when sending one after.

**Result 3:** *Messages state higher levels of work on average in the Message Before treatment compared to in the Message After treatment.*

Thus, the higher misinformation in the *MB* treatment documented in Section III.A is a result of *both* lower work and higher messages.

### *D. The Manager Decision*

The next question that naturally arises is how managers respond to messages and whether this varies by treatment. We focus on managers who received a message recommending investment. Recall that a manager is matched with multiple workers (maximum three), so most managers make three investment decisions and can potentially receive three messages.[30] We have a total of 73 subjects in the role of manager, out of which 71 subjects received at least one message recommending investment. Non-parametric tests are based on subject averages of the relevant variables.

We first explore whether managers expect work on the Joint Project to be different across the treatments. Analyzing managers' beliefs, there is no significant difference in the number of letters they think the worker will encode for the Joint

[29]In reality, workers devote 42% and 49% of their total work to the Joint Project in the *MB* and *MA* treatments, respectively ($p$-value=0.15).

[30]To make each decision independent, the managers are paid for one randomly selected decision and the worker's level of work corresponding to that decision.

Project. Overall managers expect workers to encode on average 46.3 letters in the *MB* treatment and 51.1 letters in the *MA* treatment ($p$-value=0.35). Managers' investment decisions reflect this as they are equally likely to invest across treatments. On average, managers invest 55.8 percent of the time in the *MB* treatment compared to 55.4 percent in the *MA* treatment ($n_1$=37, $n_2$=34, $p$-value=0.92).[31]

To understand how informative managers expect the messages received to be, we look at correlations between message received and the manager's expectation of the work on the Joint Project ($E^M(w_J)$). On average, managers expect messages to be more informative in *MA* ($\rho$=0.57) than in *MB* ($\rho$=0.22) ($p$=0.004). Though workers' messages are more informative in *MA* and the managers expect this directionally, we do not find this translating to managers in *MA* forming more accurate beliefs about realized work. The correlation between realized and expected work is 0.29 in *MA* and 0.25 in *MB* ($p$=0.78).

If the managers correctly discounted the messages, then on average they would have discounted by 32.1 letters in the *MB* and 17.9 letters in *MA*, the amount of actual message inflation. In our data, managers are too trusting, discounting messages less than they should. Managers in *MB* discount messages by 22 letters, thus taking into account 69% of the message inflation. In *MA*, managers discount by 9.9 letters, thus taking into account only 55% of the actual message inflation. Hence, we find that although managers correctly anticipate the higher misinformation in *MB*, they misjudge its magnitude.[32]

This has potentially meaningful consequences for the managers. Managers' expected payoffs are the same across treatments ($7.07 vs $7.24, $p$-value=0.50). However, the empirical best-response for risk-neutral managers would guarantee weakly higher payoffs for managers in *MA* than *MB* ($8.30 vs $8.60, $p$-value=0.09). Figure 5 shows the cumulative distribution functions of managers' actual expected payoffs and their expected payoffs under the assumption that work is perfectly observable. We calculate this assuming the risk-neutral best response for managers given actual worker effort. That is, if the worker encoded more than 23 letters, we say the manager will invest, but will not invest otherwise. There is a significant gap between the *MB* and *MA* best response distributions, but actual manager behavior does not capture this. Hence we find that man-

[31] Though managers saw up to three messages, we don't find any effect of decision order or message rank on investment decisions. See Appendix Table A4 for details.

[32] As further evidence, Figure A5 in the Appendix shows that the percentage of managers who incorrectly believe a message is more informative is higher in *MA* than *MB*.

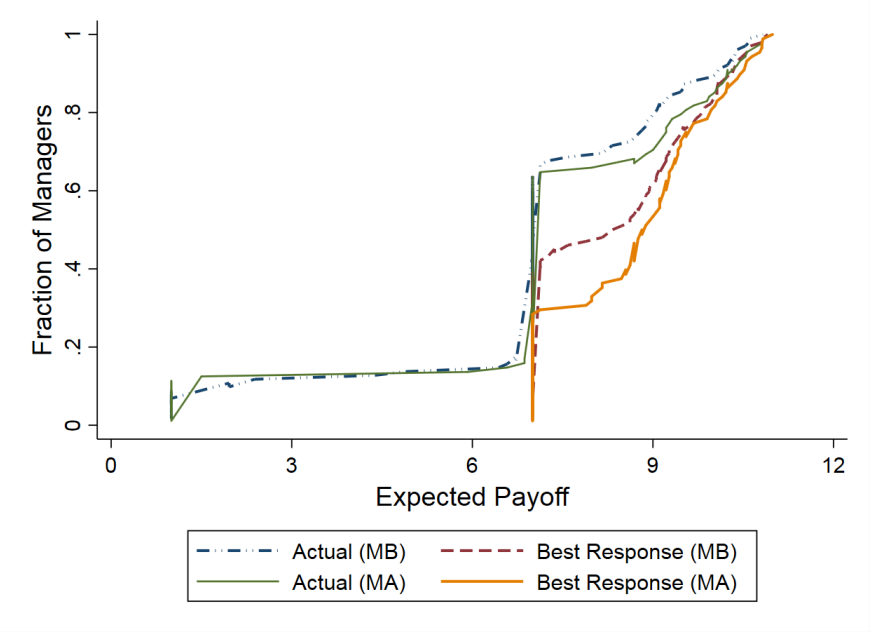



Figure 5. : CDF of Expected Payoff

*Note:* The best response is computed assuming risk-neutrality. Managers are assumed to invest only when $w_J \geq 23$.

agers in *MA* are unable to reap the potential advantages of the more informative communication.

**Result 4:** *Managers anticipate higher misinformation in Message Before compared to Message After but underestimate the extent of the treatment difference.*

## IV. Possible Explanations

Overall, we find that the timing of communication affects behavior. Workers are more honest and cooperative when communicating after taking actions compared to communicating before. We designed our experiment to be a first step in establishing the existence of a phenomenon, but our design does not allow for disentangling all possible explanations of the observed pattern. In this section, we explore possible explanations for our observed treatment differences. We provide some initial evidence for/against these explanations, but we believe focused analysis on the underlying mechanisms will be a fruitful avenue for future research.

### A. Beliefs of the Manager's Perception of the Message

It is possible that, *independent* of the workers' actual behavior, workers expect managers to discount a given message more in the *MB* treatment than in the *MA* treatment.[33] Workers could respond to such beliefs in two ways. One, workers can send a higher message in the *MB* treatment to compensate for this, leading to higher observed misinformation. Two, workers send the same message in both treatments, but work less in the *MB* treatment as they believe managers expect them to do so.[34] To test whether the difference in misinformation is a response to what workers think managers expect, we first examine whether workers expect managers to discount a given message differently across the two treatments.

Recall that we asked managers to estimate the number of letters the worker encoded for the Joint Project, $(E^M(w_J))$. Then we asked the worker to guess the number the manager reported, $(E^W(E^M(w_J)))$. We use the worker's guess as a measure of his beliefs of the amount of work the manager expects him to do. We calculate a measure of how much the worker expects the manager will discount his message by calculating the difference between the message sent and the work the worker thinks the manager expects ($Discount := m - E^W(E^M(w_J))$). We find that workers expect messages to be discounted by 17.7 letters on average in the *MB* treatment compared to 10.1 letters in the *MA* treatment ($p$-value$= 0.03$).[35]

If higher misinformation in *MB* were driven by workers' beliefs of the managers' expectations, this difference should account for the higher observed misinformation in the *MB* treatment. Table 4 augments our regressions predicting misinformation in Table 2 with the variable *Discount*. *Discount* significantly increases the amount of misinformation, although the magnitude of the effect is small. Additionally, the coefficient of *MB* is still significant implying that the difference in

[33] A possible reason could be if managers think workers' messages are unintentionally inflated in the *MB* treatment due to overconfidence.

[34] This is in line with expectations-based guilt-aversion hypotheses which proposes that individuals suffer a psychological cost proportional to the amount by which they think they fail to meet others' expectations of them (Charness and Dufwenberg, 2006; Ederer and Stremitzer, 2016; Di Bartolomeo et al., 2017). In our experiment, this would imply that a worker's effort on the Joint Project will depend on how much effort he thinks the manager expects him to put forth (the worker's second-order belief: $E^W(E^M(w_J))$). A worker will encode a higher number of letters for the Joint Project when he thinks the manager expects him to do so compared to when he thinks the manager does not expect him to do so.

[35] This is confirmed by running an OLS regression predicting workers' beliefs from the treatment dummy, the message sent and our standard set of controls. Table A5 shows that higher messages increase the number of letters the worker believes the manager expects him to encode for the Joint Project. Furthermore, controlling for the message, workers in the *MB* treatment have significantly lower second-order beliefs.

Table 4—: Amount of Misinformation

| Dependent variable: $m - w_J$ | |
|---|---|
| **Message Before** | 11.24*** |
| | (3.77) |
| **Part 1 Performance** | 0.18 |
| | (0.21) |
| **Overestimate** | -0.39 |
| | (0.30) |
| **Overestimate*Message Before** | 1.08** |
| | (0.45) |
| **Discount** | 0.65*** |
| | (0.06) |
| | (12.60) |
| Controls | Yes |
| No. of Obs. | 190 |
| R-Squared | 0.39 |

*Note:* * $p$ <0.10, ** $p$ <0.05, *** $p$ <0.01. Numbers in parentheses are robust standard errors. Controls include a gender dummy, a dummy for native language, and year-of-study dummies. 3 observations were dropped since being a graduate student perfectly predicts investment. Message Before is a treatment dummy variable (1=*MB*, 0=*MA*), Part 1 Performance is a subject's performance in the Part 1 calibration, Overestimate is a measure of forecast error, and Discount is how much the worker expects the manager will discount his message.

expectations is not sufficient to explain the difference in misinformation between the two treatments. Thus, it's not the case that the difference in misinformation across treatments is due solely to workers' differing expectations of the managers.

### *B. Moral Wiggle Room*

A second potential explanation of our treatment differences is that workers' intentions are less apparent in the *MB* treatment and the workers may use this moral wiggle room or veil of deniability to be more dishonest. This is perfectly captured in one subject's post-session questionnaire justification of why he would prefer to be the manager in *MA* rather than *MB*:

- "*I think (workers) may use the failed promise as an excuse to encode more for themselves and just say they couldn't do as much as they hoped*"

In our game, managers are unable to infer whether a failed Joint Project in the *MB* treatment was a result of intentional deception or was due to the worker overestimating his ability and not being able to encode as many letters as expected. This concern is not present in *MA* since workers communicate after seeing their realized effort. If workers have a preference for *appearing* truthful rather than

actually being truthful, they may exploit this second-order uncertainty to deliberately inflate their message in the *MB* treatment.

We use standard one-shot matrix games to address the impact of moral wiggle room and explore the robustness of our results to other decision contexts. The binary nature of the task eliminates concerns about managers' inability to infer workers' intentions, so moral wiggle room has no room to affect behavior in these games. After the real-effort task, before receiving any feedback on Part 2, participants made decisions in five 2×2 matrix games. In each game, the sender first chooses an action, followed by the receiver. Even though the sender moves first, his choice is unobservable to the receiver. As in the real-effort task, the sender can signal his action to the receiver through a message. Treatments vary in when the sender could send the message, either before or after he took his action. In the message, the sender could signal his intended action or chosen action depending upon the treatment, as well as recommend an action to the receiver.

Our primary interest is in the games depicted in Table 5. Note, the games in Table 5 have the same strategic considerations as the real-effort manager-worker game, with a reduction in the number of choices available for the worker (in this case the sender). The sender has a dominant strategy to choose *D*, identical to the worker having a dominant strategy to work on his Personal Project in the real-effort task. The receiver faces a coordination game, where she wants to choose *C* if the sender chooses *C* (invest if worker works on Joint Project), else choose *D* (not invest). Assuming self-interested players with no costs of deception, the Nash equilibrium outcome for the game with communication is both players choose *D*. However, if individuals are other-regarding and/or suffer costs of deceiving, we would expect outcomes to be more cooperative as seen in the real-effort task. We look to see whether subjects misinform more in the *MB* treatment than in the *MA* treatment, in line with results from our real-effort task.

Table 5—: Matrix Games in the *Choice Task*

| | | **Receiver** | |
|---|---|---|---|
| | | *D* | *C* |
| **Sender** | *D* | 70, 70 | 130, 30 |
| | *C* | 30,80 | 90, 90 |

(a) Manager-Worker (High)

| | | **Receiver** | |
|---|---|---|---|
| | | *D* | *C* |
| **Sender** | *D* | 70, 70 | 110, 30 |
| | *C* | 30,80 | 90, 90 |

(b) Manager-Worker (Low)

In addition, we present subjects with two versions of this game to directly test how behavior responds to changing the benefit from misinformation. The games in Table 5 differ only in that the temptation payoff for the sender is reduced from 130 (high stakes) to 110 (low stakes). This weakens the incentives to misinform and, if individuals have a positive cost of deceiving, we should see less frequent deception in the low stake games as compared to the high stake games.

At the beginning of the Choice Task, participants were randomized into roles of sender and receiver and played in fixed roles for all five rounds. In each round, they were presented with a different payoff matrix. In addition to the two games in Table 5, we include three other games for robustness. Discussion of these games can be found in Appendix Section B.B4. The order of the games was randomized across sessions. If this part were chosen for payment, participants were paid for their decisions in one randomly selected round.

Table 6—: Signal and Actions

| | **High Stakes** | | | **Low Stakes** | | |
|---|---|---|---|---|---|---|
| | Message Before | Message After | $p$-value | Message Before | Message After | $p$-value |
| Percent Misinforming | 63.5 | 38.2 | 0.003 | 45.9 | 48.5 | 0.75 |
| Percent Signaling C | 91.9 | 75.0 | 0.006 | 82.4 | 80.9 | 0.81 |
| Percent Choosing C | 29.7 | 44.1 | 0.07 | 36.5 | 32.2 | 0.61 |

Results. — We begin by reporting the fraction of senders who misinform.[36] The first row of Table 6 indicates the fraction of senders who misinform in each game across treatments. In the high stakes game, we find strong confirmation of our previous results—senders deceive significantly more in the *MB* than the *MA* treatment. However, the data fail to support these hypotheses when considering the low stake games.

These results are formalized in Table 7 which reports results from probit regressions predicting whether or not a sender misinformed. In addition to the treatment dummy (*Message Before*) and demographic controls, we include (i) *Round*, a variable signifying the period in which the game was presented, and (ii)

[36] Although the sender can misinform the receiver in two possible ways - by signaling $C$ while choosing $D$ or by signaling $D$ while choosing $C$ - the latter strategy is hard to rationalize and extremely rare, constituting less than 1% of our observations. Our main analysis includes these observations, although all results are robust to excluding them.

*Worker*, a dummy equal to one if the the sender had been a worker in the real-effort task. The positive coefficient on *Message Before* in Column (1) confirms our conjecture that individuals are more reluctant to lie about a past action as compared to a future action in the high stakes game. Being in the *MB* treatment increases the probability of misinforming by 24 percentage points. Since treatments no longer differ in how transparent the sender's intentions are, this difference cannot be driven by moral wiggle room.

Table 7—: Probit predicting whether sender misinforms in the Choice Task

| Dependent variable: Probability Sender Misinforms | | |
|---|---|---|
| | Manager-Worker | |
| | High | Low |
| | (1) | (2) |
| **Message Before** | 0.24*** | 0.04 |
| | (0.07) | (0.08) |
| **Round** | 0.02 | 0.05** |
| | (0.03) | (0.03) |
| **Worker** | -0.01 | 0.06 |
| | (0.10) | (0.10) |
| Controls | Yes | Yes |
| No. of Obs. | 142 | 142 |

*Note:* * $p$ <0.10, ** $p$ <0.05, *** $p$ <0.01. Numbers in parenthesis are robust standard errors. Controls include a dummy for gender, a dummy if the participant's native language is English, dummies for year in school, and a dummy for being a graduate student. Message Before is a treatment dummy variable (1=*MB*, 0=*MA*), Worker is a dummy variable for role in Part 2 (1=worker, 0=manager), and Round is the order of the matrix game. Marginal effects reported.

In the low stake games, while we observe a positive coefficient on *Message Before*, this is not statistically significant. There are several possible explanations for this. It is possible that the timing is relevant only for a subset of participants—those who have substantial costs of misinforming. Since the benefit from misinforming shrinks in the low-stake game (the difference between the temptation payoff and cooperation payoff is only \$2), one would expect only senders with very low costs of deception to misinform. Alternatively, note that the sender's decision and signal are binary variables. Unlike the real-effort task, where we captured the *size* of the misinformation, the choice task only allows us to capture whether or not the sender misinformed. This could possibly make it harder to identify a small treatment effect.

Analogous to the real-effort task, we can decompose the misinformation in the Choice Task into its two components—signal and action. Due to the binary

nature of our decision variables, we compare the frequency of cooperative signals and actions across treatments. Table 6 displays the overall fraction of senders who signal the cooperative action $C$ and the fraction of senders who choose $C$. Recall that, in the real-effort task, messages sent before the action was taken promised a higher level of cooperation. Row 2 of Table 6 indicates that same pattern. A vast majority of senders choose to signal $C$ in both treatments. The percent of $C$ signals is always higher in the $MB$ treatment, and significantly more so in the High stakes game. Not only do workers send higher messages in $MB$, but they cooperate less often. Row 3 shows that individuals are significantly more likely to choose $C$ in $MA$ than in $MB$ in the High stakes game.

Overall, we find that individuals misinform more in $MB$ than $MA$ when the potential gains from misinformation are high, even when we remove potential moral wiggle room. This suggests that moral wiggle room is not the only mechanism driving our main result.

### *C. Differential Cost of Deception*

We asked subjects directly whether and why they think that timing of communication would impact behavior. In a post-session questionnaire all subjects (managers and workers) were presented with the following question before seeing the results:

> *Imagine you are (the manager) and you can now determine when you want (the worker) to send you the message. Choose which scenario you would pick.*
> *Scenario 1: (The worker) sends a message before he begins working.*
> *Scenario 2: (The worker) sends a message after he finishes working.*
> *In both scenarios, you do not learn how many letters (the worker) encoded or whether or not the project was successful. You only see his message before deciding to invest.*

Table 8 presents their responses. A large percentage of subjects state a strict preference for the worker sending a message only after he has worked. We also asked subjects to give reasons supporting their choice of response. Though the reasons varied, they could be categorized into three broad justifications—accuracy, differential cost of deception, and moral wiggle room.

Table 8—: Preferred Treatment

| *Responses* | *MB* | *MA* | *p*-value |
|---|---|---|---|
| Prefer MB | 10.8 | 21.0 | 0.02 |
| Prefer MA | 51.3 | 44.8 | 0.27 |
| Indifferent | 34.6 | 37.8 | 0.57 |

*Notes:* Percent of workers in each treatment who stated preferring *MB*, *MA* or being indifferent to the two treatments.

The most commonly cited argument ($\approx 60\%$) for preferring Scenario 2 was that the worker could have a better estimate of his performance when sending a message after working. Some subjects qualified this argument by stating this preference was based on the assumption that workers would be honest. However, given our analysis we already know this uncertain ability is not the reason driving the difference in behavior. Some representative responses are given below.

- "*At least there's a better chance of the number being more accurate than a projection before Player A starts.*"
- "*Would prefer to know what was actually done, assuming they are being honest.*"

The remaining justification, and the explanation we find most compelling, is that the moral cost of deception in *MA* is higher than that in *MB*. A number of participants' responses hinted at it being psychologically more difficult to deceive later ($\approx 27\%$), with a few illustrative responses below.

- "*I think it's harder to lie about something you *just* did.*"
- "*It would be harder to lie, knowing what the results were.*"
- "*...After they've already completed it, there is no uncertainty, and lying about the number would weigh more heavily on their conscience.*"

These responses point towards the idea that a realized level of work is morally difficult to misrepresent outright. It may be that subjects find it harder to attribute their dishonesty to external factors when the work has been realized.[37]

[37]We also asked participants who they would punish more: a person who broke his promise or a liar. 71% of participants stated wanting to punish a liar more than someone who broke his promise.

## V. Discussion

Over the past decade, an extensive literature has documented that non-binding statements of intent or promises can be informative and can increase cooperation in social dilemmas. But not all instances of communication are forward-looking. We design a two-player hidden action game to compare ex-ante and ex-post communication. We find that a communication regime where individuals report on past decisions results in more truthful communication and in higher overall effort than one in which individuals communicate about their intended future effort. Our results show that timing of communication is a critical variable that merits attention in the design of mechanisms. In addition, our results raise questions about the appropriateness of conflating promise-breaking and lying about a past action. Over the years, legal thinkers and philosophers have discussed whether misrepresenting intent is the same as misrepresenting a fact (Cavico, 1997; Ayres and Klass, 2008). We provide empirical investigation into this question and show that they are, in fact, different.

We hypothesize that the observed difference in our data has two behavioral foundations. First, the moral cost of lying about a past action is higher than the cost of breaking a promise. Qualitative responses from subjects, as well as our own introspection, suggest that lies about past actions weigh more heavily on our consciences than do broken promises. Previous papers have shown that higher mutability of an outcome is associated with more misreporting (Batson et al., 1997, Shalvi et al., 2011, Shalvi, Eldar and Bereby-Meyer, 2012, Shalvi et al., 2015). We don't directly manipulate mutability, which was the focus of their papers, but "moral wiggle room" under ex-ante communication could be thought to play a similar role. The future is inherently more mutable than the past, which could contribute to higher misinformation in statements about future actions. We think it would be interesting for future research to look into disentangling the pure timing aspect from other related notions of mutability and uncertainty.

Second, our results on the patterns of work over time suggest that communication may trigger moral responses that operate, at least partially, through salience which is asymmetric for past and future actions. Such temporal asymmetry has previously been demonstrated in Caruso, Gilbert and Wilson (2008) in a non-strategic decision context. Caruso, Gilbert and Wilson (2008) find that individuals value future events more than past events. Whether this "temporal

value asymmetry" is similarly driving results in our environment leaves an interesting question for future research. In addition, we know relatively little about the impact of communication in environments with long time horizons. Our paper takes a first step in addressing this, but our understanding of communication would benefit greatly from more focused study in this area.

Finally, our results bring to light the need for more concentrated research on the salience effect of norms and moral decision making. Shu et al. (2012) find that signing tax forms on top versus on the bottom increases the frequency of truthful reporting. They suggest that signing on top primes individuals to have morality at the top of mind, so they are more likely to follow the honest social norm. In their environment, however, individuals are deciding how truthfully to report on an exogenous outcome.[38] We show that when a person is able to jointly optimize message and action, the patterns of lying may be reversed. Our environment differs from theirs in many other regards. In particular, subjects were primed with a moral stance to report truthfully in their experiment. In contrast, we made no mention of morality and subjects were free to choose their own honesty levels. Most daily interactions are free from moral priming, so if this aspect of the environment is contributing to the differences in our results, the direction of misreporting in our paper is what we should expect to observe more frequently. Experiments looking into the aspects of these environments that contribute to these differences in behavior would be very interesting. This strikes us as particularly important for policy makers and institutional designers who may wish to use this information to nudge behavior toward truth-telling and cooperation.

## REFERENCES

**Abeler, Johannes, Anke Becker, and Armin Falk.** 2014. "Representative evidence on lying costs." *Journal of Public Economics*, 113: 96–104.

**Arya, Anil, John Fellingham, Jonathan Glover, and Kashi Sivaramakrishnan.** 2000. "Capital budgeting, the hold-up problem, and information system design." *Management Science*, 46(2): 205–216.

[38] In their paper, subjects were not aware of the reporting opportunity when completing the task about which they later reported. As a result, their message became akin to reporting on an exogenous state as the outcome had already been determined.

**Ayres, Ian, and Gregory Klass.** 2008. *Insincere promises: The law of misrepresented intent.* Yale University Press.

**Batson, C. Daniel, Diane Kobrynowicz, Jessica L. Dinnerstein, and Angela D. Wilson.** 1997. "In a very different voice: Unmasking moral hypocrisy." *Journal of Personality and Social Psychology*, 1335–1348.

**Battigalli, Pierpaolo, Gary Charness, and Martin Dufwenberg.** 2013. "Deception: The role of guilt." *Journal of Economic Behavior & Organization*, 93: 227–232.

**Brandts, Jordi, Matthew Ellman, and Gary Charness.** 2016. "Let's talk: How communication affects contract design." *Journal of the European Economic Association*, 14(4): 943–974.

**Brosig, Jeannette, Magdalena Margreiter, and Joachim Weimann.** 2005. *Endogenous group formation and the provision of public goods: The role of promises and lies.* Univ., FEMM.

**Cai, Hongbin, and Joseph Tao-Yi Wang.** 2006. "Overcommunication in strategic information transmission games." *Games and Economic Behavior*, 56(1): 7–36.

**Caruso, Eugene M, Daniel T Gilbert, and Timothy D Wilson.** 2008. "A wrinkle in time asymmetric valuation of past and future events." *Psychological Science*, 19(8): 796–801.

**Cavico, Frank J.** 1997. "Fraudulent, Negligent, and Innocent Misrepresentation in the Employment Context: The Deceitful, Careless, and Thoughtless Employer." *Campbell L. Rev.*, 20: 1.

**Charness, Gary.** 2000. "Self-serving cheap talk: A test of Aumann's conjecture." *Games and Economic Behavior*, 33(2): 177–194.

**Charness, Gary, and Martin Dufwenberg.** 2006. "Promises and partnership." *Econometrica*, 74(6): 1579–1601.

**Charness, Gary, and Martin Dufwenberg.** 2010. "Bare promises: An experiment." *Economics Letters*, 107(2): 281–283.

**Charness, Gary, David Masclet, and Marie Claire Villeval.** 2014. "The dark side of competition for status." *Management Science*, 60(1): 38–55.

**Church, Bryan K, R Lynn Hannan, and Xi Jason Kuang.** 2012. "Shared interest and honesty in budget reporting." *Accounting, Organizations and Society*, 37(3): 155–167.

**Clark, Jeremy, and Lana Friesen.** 2009. "Overconfidence in forecasts of own performance: An experimental study." *The Economic Journal*, 119(534): 229–251.

**Clark, Kenneth, Stephen Kay, and Martin Sefton.** 2001. "When are Nash equilibria self-enforcing? An experimental analysis." *International Journal of Game Theory*, 29(4): 495–515.

**Cohen, Elaine.** 2011. "Inflating Community Impacts." *online at http://csr-reporting.blogspot.com/2011/04/inflating-community-impacts.html*.

**Cooper, Russell, Douglas V DeJong, Robert Forsythe, and Thomas W Ross.** 1992. "Communication in coordination games." *The Quarterly Journal of Economics*, 107(2): 739–771.

**Crawford, Vincent P, and Joel Sobel.** 1982. "Strategic information transmission." *Econometrica: Journal of the Econometric Society*, 1431–1451.

**Desai, Sreedhari D, and Maryam Kouchaki.** 2015. "Work-report formats and overbilling: how unit-reporting vs. cost-reporting increases accountability and decreases overbilling." *Organizational Behavior and Human Decision Processes*, 130: 79–88.

**Di Bartolomeo, Giovanni, Martin Dufwenberg, Stefano Papaa, and Francesco Passarelli.** 2017. "Promises, Expectations & Causation." *Sapienza University of Rome Working Paper*.

**Ederer, Florian, and Alexander Stremitzer.** 2016. "Promises and expectations."

**Erkal, Nisvan, Lata Gangadharan, and Nikos Nikiforakis.** 2011. "Relative earnings and giving in a real-effort experiment." *The American Economic Review*, 101(7): 3330–3348.

**Farrell, Joseph.** 1988. "Communication, coordination and Nash equilibrium." *Economics Letters*, 27(3): 209–214.

**Fellingham, John C, and Richard A Young.** 1990. "The value of self-reported costs in repeated investment decisions." *Accounting Review*, 837–856.

**Fischbacher, Urs.** 2007. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental economics*, 10(2): 171–178.

**Fischbacher, Urs, and Franziska Föllmi-Heusi.** 2013. "Lies in disguise - an experimental study on cheating." *Journal of the European Economic Association*, 11(3): 525–547.

**Glaeser, Edward L., David I. Laibson, Jose A. Scheinkman, and Christine L. Soutter.** 2000. "Measuring trust." *Quarterly Journal of Economics*, 811–846.

**Gneezy, Uri.** 2005. "Deception: The role of consequences." *American Economic Review*, 384–394.

**Greiner, Ben.** 2004. "An Online Recruitment System for Economics Experiments."

**Khalmetski, Kiryl, and Gilad Tirosh.** 2012. "Two types of lies under different communication regimes."

**Lundquist, Tobias, Tore Ellingsen, Erik Gribbe, and Magnus Johannesson.** 2009. "The aversion to lying." *Journal of Economic Behavior & Organization*, 70(1): 81–92.

**Mazar, Nina, On Amir, and Dan Ariely.** 2008. "The dishonesty of honest people: A theory of self-concept maintenance." *Journal of marketing research*, 45(6): 633–644.

**Miettinen, Topi, and Sigrid Suetens.** 2008. "Communication and Guilt in a Prisoner's Dilemma." *Journal of Conflict Resolution*, 52(6): 945–960.

**Rothbart, Myron, and Mark Snyder.** 1970. "Confidence in the prediction and postdiction of an uncertain outcome." *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 2(1): 38.

**Sánchez-Pagés, Santiago, and Marc Vorsatz.** 2007. "An experimental study of truth-telling in a sender–receiver game." *Games and Economic Behavior*, 61(1): 86–112.

**Schlag, Karl H, and Péter Vida.** 2015. "Believing when Credible: Talking about Future Plans and Past Actions."

**Schotter, Andrew, and Isabel Trevino.** 2014. "Belief elicitation in the laboratory." *Annu. Rev. Econ.*, 6(1): 103–128.

**Serra-Garcia, Marta, Eric Van Damme, and Jan Potters.** 2013. "Lying about what you know or about what you do?" *Journal of the European Economic Association*, 11(5): 1204–1229.

**Shalvi, Shaul, Francesca Gino, Rachel Barkan, and Sahar Ayal.** 2015. "Self-serving justifications: Doing wrong and feeling moral." *Current Directions in Psychological Science*, 125–130.

**Shalvi, Shaul, Jason Dana, Michel J.J. Handgraaf, and Carsten K.W. De Dreu.** 2011. "Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior." *Organizational Behavior and Human Decision Processes*, 181–190.

**Shalvi, Shaul, Ori Eldar, and Yoella Bereby-Meyer.** 2012. "Honesty requires time (and lack of justifications)." *Psychological Science*, 1264–1270.

**Shu, Lisa L, Nina Mazar, Francesca Gino, Dan Ariely, and Max H Bazerman.** 2012. "Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end." *Proceedings of the National Academy of Sciences*, 109(38): 15197–15200.

**Vanberg, Christoph.** 2008. "Why do people keep their promises? an experimental test of two explanations." *Econometrica*, 76(6): 1467–1480.

**Van den Assem, Martijn J, Dennie Van Dolder, and Richard H Thaler.** 2012. "Split or steal? Cooperative behavior when the stakes are large." *Management Science*, 58(1): 2–20.

**Ward, Jennifer Inez.** 2014. "Missed Targets: When Companies Fail to Keep Their Key Sustainability

Promises." *online at https://www.theguardian.com/sustainable-business/blog/2014/jul/21/sustainability-goals-promise-broken-failure-target-walmart-disney.*

**Zultan, Ro'i.** 2013. "Timing of messages and the Aumann conjecture: a multiple-selves approach." *International Journal of Game Theory*, 42: 789–800.

# Appendix A: Additional Tables and Figures

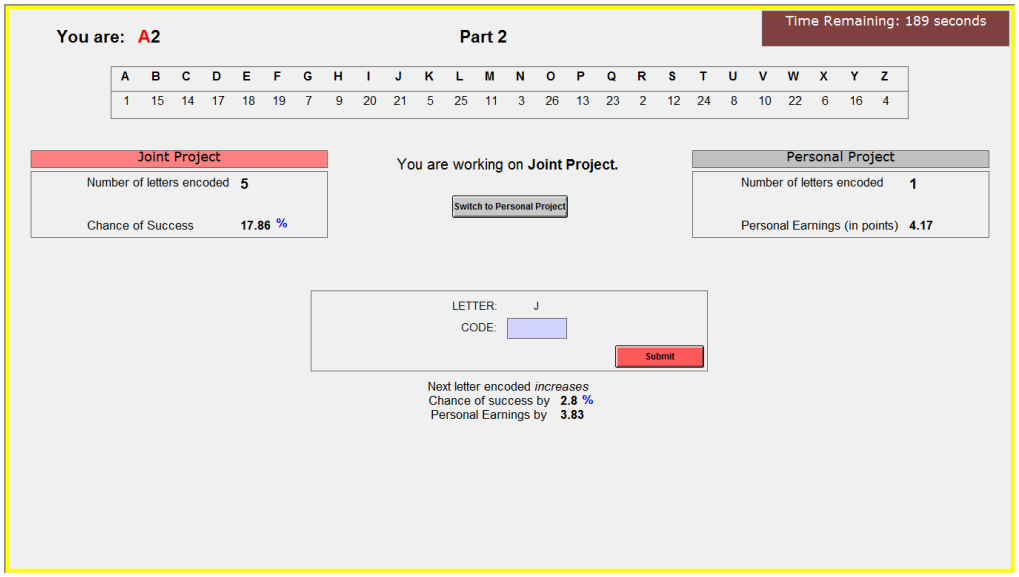



Figure A1. : Work Stage Screen

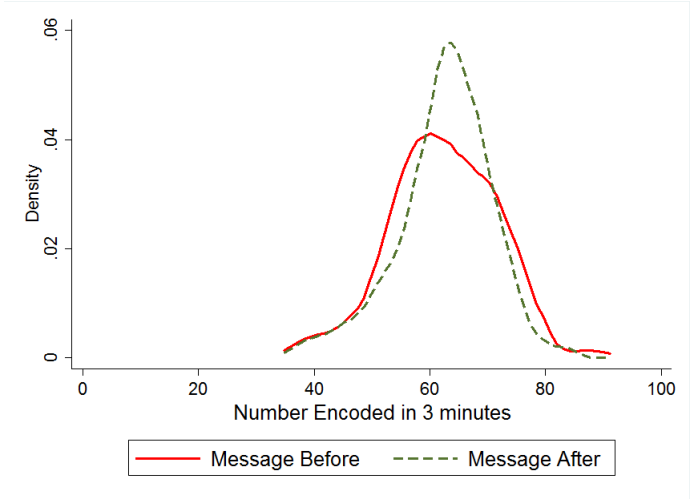



Figure A2. : Ability Distribution in the Effort Task

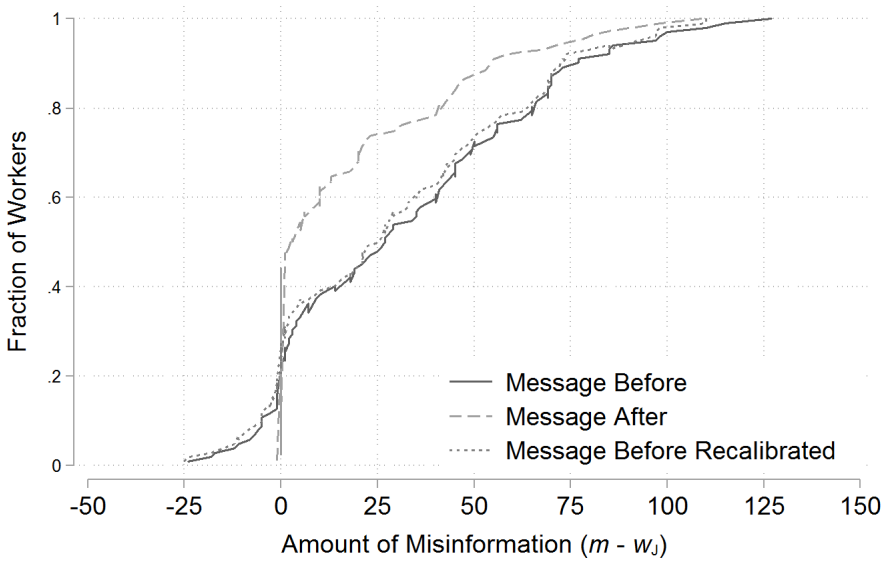



Figure A3. : Distribution of Misinformation

Table A1—: Estimated and Actual Mean number of letters encoded in Part 2

| | **Message Before** | **Message After** | $p$-value |
|---|---|---|---|
| Estimate ($\hat{w}_{total}$) | 85.39 | 85.44 | 0.98 |
| | (12.3) | (12.5) | |
| Actual ($w_{total}$) | 87.6 | 86.8 | 0.65 |
| | (12.9) | (11.5) | |
| Overestimate | -2.25 | -1.4 | 0.51 |
| ($\hat{w}_{total} - w_{total}$) | (9.4) | (7.8) | |

*Note:* Standard deviations reported in parenthesis. $p$-values are calculated using Fisher-Pitman permutation test using Monte Carlo method with 200,000 simulations.

Table A2—: Effort Allocation in the Joint Project

| Dependent Variable: | Fraction of total work allocated to the Joint Project ($\frac{w_J}{w_{total}}$) | Fraction of total time spent working on the Joint Project ($\frac{t_J}{240}$) | Probability of Success of the Joint Project ($p = \frac{w_J}{w_J+23} \times 100$) |
|---|---|---|---|
| **Message Before** | -0.08* | -0.08* | -7.44* |
| | (0.05) | (0.05) | (4.04) |
| **Part 1 Performance** | -0.00 | -0.00 | 0.23 |
| | (0.00) | (0.00) | (0.26) |
| **Constant** | 0.65*** | 0.67*** | 44.38** |
| | (0.21) | (0.20) | (17.63) |
| Controls | Yes | Yes | Yes |
| No. of Obs. | 190 | 190 | 190 |
| R-Squared | 0.07 | 0.08 | 0.08 |

*Note:* * $p$ <0.10, ** $p$ <0.05, *** $p$ <0.01. Numbers in parenthesis are robust standard errors. Controls include dummy if the participant's native language is English, dummies for year in school and a dummy for graduate student. Message Before is a treatment dummy variable (1=*MB*, 0=*MA*) and Part 1 Performance is the subject's performance in the Part 1 calibration.

Table A3—: Messages

| Dependent variable: Content of Message | | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **Message Before** | 7.38** | 6.45** | 7.39** | 9.78*** |
| | (3.29) | (3.26) | (3.29) | (3.04) |
| **Part 1 Performance** | | 0.60*** | 0.65*** | 0.56*** |
| | | (0.18) | (0.18) | (0.17) |
| **Overestimate** | | | 0.12 | -0.01 |
| | | | (0.42) | (0.35) |
| **Overestimate*Message Before** | | | 0.84 | 1.15** |
| | | | (0.60) | (0.56) |
| **Number of Letters Encoded for Joint Project** | | | | 0.27*** |
| | | | | (0.05) |
| **Constant** | 61.16*** | 24.01* | 19.53 | 12.13 |
| | (2.41) | (13.24) | (13.22) | (11.68) |
| Controls | No | Yes | Yes | Yes |
| No. of Obs. | 190 | 190 | 190 | 190 |
| R-Squared | 0.03 | 0.11 | 0.14 | 0.26 |

* $p$ <0.10, ** $p$ <0.05, *** $p$ <0.01. Numbers in parenthesis are robust standard errors. Controls include a dummy for gender, a dummy for native language, year-of-study dummies and a dummy for being a graduate student. Message Before is a treatment dummy variable (1=*MB*, 0=*MA*), Part 1 Performance is the subject's performance in the Part 1 calibration, and Overestimate is a measure of forecast error.

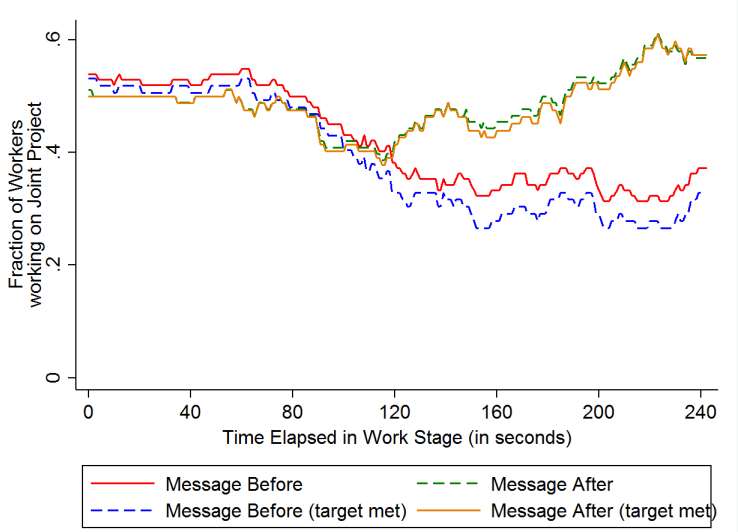



Figure A4. : Distribution of Work Over Time for Workers Who Could Achieve Stated Message

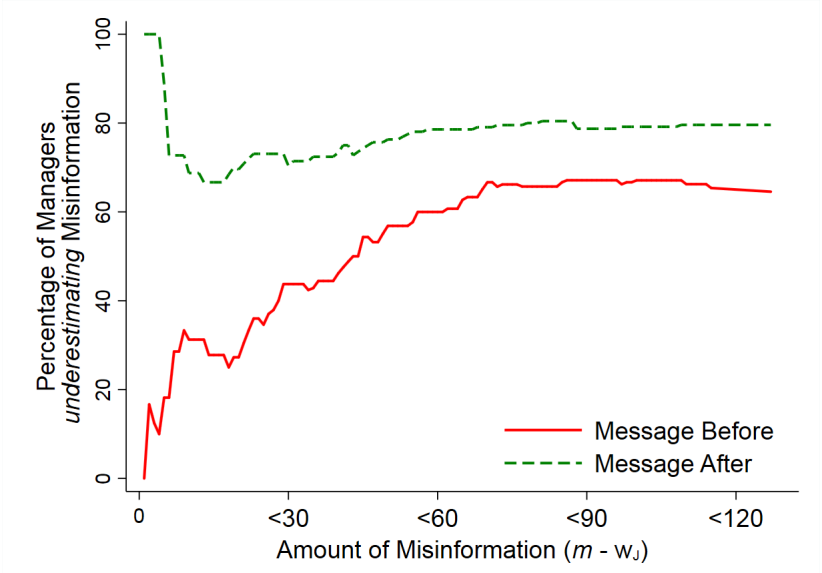



Figure A5. : Fraction of manager decisions where the manager underestimates the actual level of message inflation

*Note:* The data include all observation where the message was inflated. The x-axis depicts The y-axis depicts the percentage of manager decisions where the managers underestimates the extent to which the message received is inflated i.e. $m - E^M(w_J) < m - w_J$.

Table A4—: Effect of decision order and message rank on investment

| | Dependent variable: Investment | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Message Before | -0.06 | -0.06 | -0.06 |
| | (0.07) | (0.07) | (0.07) |
| Mesage | 0.04*** | 0.04*** | 0.03*** |
| | (0.01) | (0.01) | (0.01) |
| Message$^2$ | -0.00*** | -0.00*** | -0.00*** |
| | (0.00) | (0.00) | (0.00) |
| Part 1 Performance | -0.01 | -0.01 | -0.01 |
| | (0.01) | (0.00) | (0.01) |
| Decision Order | | -0.03 | |
| | | (0.04) | |
| Message Rank | | | 0.05 |
| | | | (0.05) |
| Clusters | 70 | 70 | 70 |
| No. of Obs. | 187 | 187 | 187 |

* $p$ <0.10, ** $p$ <0.05, *** $p$ <0.01. Standard errors are clustered at the subject level. Controls include a dummy for gender, a dummy for native language, and year-of-studies dummies. Message Before is a treatment dummy variable (1=*MB*, 0=*MA*), Message is the message received by the manager, Message$^2$ squares this message due to the concave probability function, Part 1 Performance is the subject's performance in the Part 1 calibration, Decision Order is the order in which the manager saw the message (from top of screen to bottom), and Message Rank is the relative ranking of the message content among all messages the manager received. 3 observations are dropped as being a graduate student perfectly predicts investment.

Table A5—: Predicting workers' Second-Order Beliefs

| Dependent variable: $Belief_{worker}$ | |
|---|---|
| **Message Before** | -7.31** |
| | (3.16) |
| **Message** | 0.83*** |
| | (0.07) |
| **Part 1 Performance** | -0.03 |
| | (0.22) |
| Controls | Yes |
| No. of Obs. | 190 |
| R-Squared | 0.42 |

* $p$ <0.10, ** $p$ <0.05, *** $p$ <0.01. Numbers in parenthesis are robust standard errors. Controls include a gender dummy, a dummy for native language, year-of-study dummies and a dummy for being a graduate student. Message Before is a treatment dummy variable (1=$MB$, 0=$MA$, Message is the message sent by the worker, and Part 1 Performance is the subject's performance in the Part 1 calibration.

## Appendix B: Additional Results

### *B1. Conservative Messages*

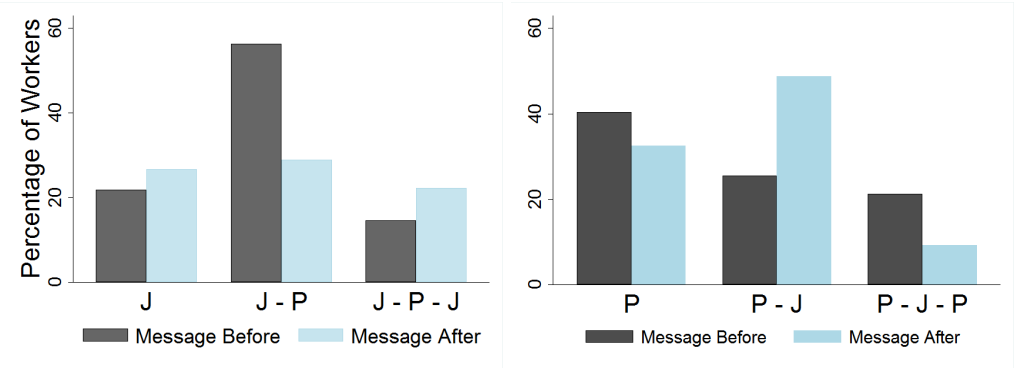
Though the frequency of inflated messages is higher in the *MB* treatment as compared to the *MA* treatment, there is a positive fraction of workers (16 subjects) whose messages are conservative in the *MB* treatment. This can be seen from the negative starting point of the CDF of misinformation in the *MB* treatment. Since it is not strategically beneficial to the worker to under-promise, one potential explanation can be that these workers mistakenly exceed the number promised. This is possible if subjects fail to notice the number of current encoded letters on their screens and forget to make the switch to working on the Personal Project. In fact, 11 out of these 16 workers with conservative messages started off working for the Joint Project and subsequently switched to the Personal Project. 25% switched after exceeding the promised amount by 1 letter, 50% by 5 letters, and 100% within 8 letters, suggesting that conservative messages may be driven partly by inattention. 4 workers started working on the Joint Project but never switched. Most likely, these are workers who underestimated how much they could encode in total; they wanted to work only on the Joint Project, so they simply keep working even after realizing they can encode more than promised. It is important to note that since misinformation is higher in the *MB* treatment, such conservative messages only bring down the average message inflation in the *MB* treatment. We find that if we treat conservative messages as fully informative (i.e. counted as zero rather than negative misinformation), average misinformation is 85% higher in *MB* than in *MA* ($p$ <0.001, $n$=190).

### *B2. Switching Patterns*

Overall across both treatments, the most frequent strategy is to switch once (42% in *MB* and 39% in *MA*). However, switching patterns are most informative when we condition on the project the worker begins with since a single switch from Joint to Personal reflects moving towards cooperation while a switch from Personal to Joint reflects moving away from cooperation.

Figure B1 depicts the three most frequent switching patterns conditional on initial choice of project.[39] When workers start with the Joint Project (Figure

[39]Only 12% of workers switch more than twice.




(a) Start with Joint Project (b) Start with Personal Project

*Note:* J: worked on Joint Project for all four minutes. P: worked on Personal Project for all four minutes. J-P: started with the Joint Project and switched over to the Personal Project. P-J: started with the Personal Project and switched over to the Joint Project. J-P-J: started with the Joint Project, switched over to the Personal Project, and switched back to the Joint Project. P-J-P: started with the Personal Project, switched over to the Joint Project, and switched back to the Personal Project.

Figure B1. : Switching Patterns

B1a), a higher frequency of workers switch once to the Personal Project in *MB* than in *MA* (56% vs 28%, $p$-value=0.006), as shown in the middle two bars. On the other hand, when workers start with the Personal Project (Figure B1b), a lower frequency of workers switch to the Joint Project in *MB* than in *MA* (25% vs 48%, $p$-value=0.02). Hence, we find a general tendency of workers moving away from the Joint Project at higher rates in *MB* than in *MA*. We find similar patterns when we look at zero (bars 1 and 2) and two (bars 3 and 4) switches. These results suggest that the divergence in cooperative behavior observed over time is not driven by a few select workers but is more widespread.

### *B3. Evidence for Timing of Communication Affecting Work Decisions for Less Altruistic Individuals*

To expand the analysis on work decision, Figure B2 shows the cumulative number of letters encoded for the Joint Project. A Kolmogorov-Smirnov test shows no difference between the two distributions ($p$-value=0.15). However, the Message After treatment stochastically dominates the Message Before treatment, with the largest differences occurring for low levels of work. As we observed in Table 3 in the main text, the difference in aggregate work stems from the work being different in the lower quartiles.

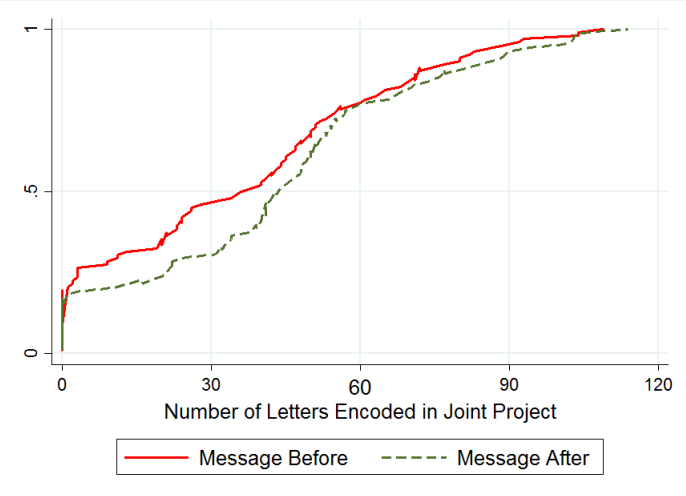



Figure B2. : Distribution of Work in Joint Project

Why does timing affect only the lower quantiles of the distribution of work done for the Joint Project? Since the lower quantiles represent workers who are encoding fewer letters for the Joint Project, its natural to conjecture that they are individuals who are less altruistic.[40] To be clear, by altruism we refer to a baseline preference for cooperation or other-regarding behavior, in the absence of communication.[41] For a subset of the sample we have a coarse measure of altruism. We will use this measure (described below) to see if altruism measures correlate with work done on the Joint Project.

In Sessions 10-16, at the end of the experiment we collected data on a series of binary decisions where subjects chose between two options - Option A, which gives both him and another random player 90 points each, and Option B, which pays the subject some positive payoff $x$ and the other player 30 points. Subjects made 13 decisions, presented to them in a list (Figure B3) which varied the value of $x$ from 90 to 150 points in intervals of 5 points. We use the subject's switch point (the row he starts preferring Option B to Option A) as a measure for his altruistic preferences. Switching in later rows indicate a higher level of altruism.

In the sub-sample of workers for whom we have data on this measure, we find

[40] Alternatively, since working on the Joint Project is also risky for the worker, as he is unsure of the manager's investment, these may be individuals who are risk-averse.

[41] This is not to be confused with a cooperative *outcome*, which may be the result of altruistic preferences or (and) positive costs of deception.

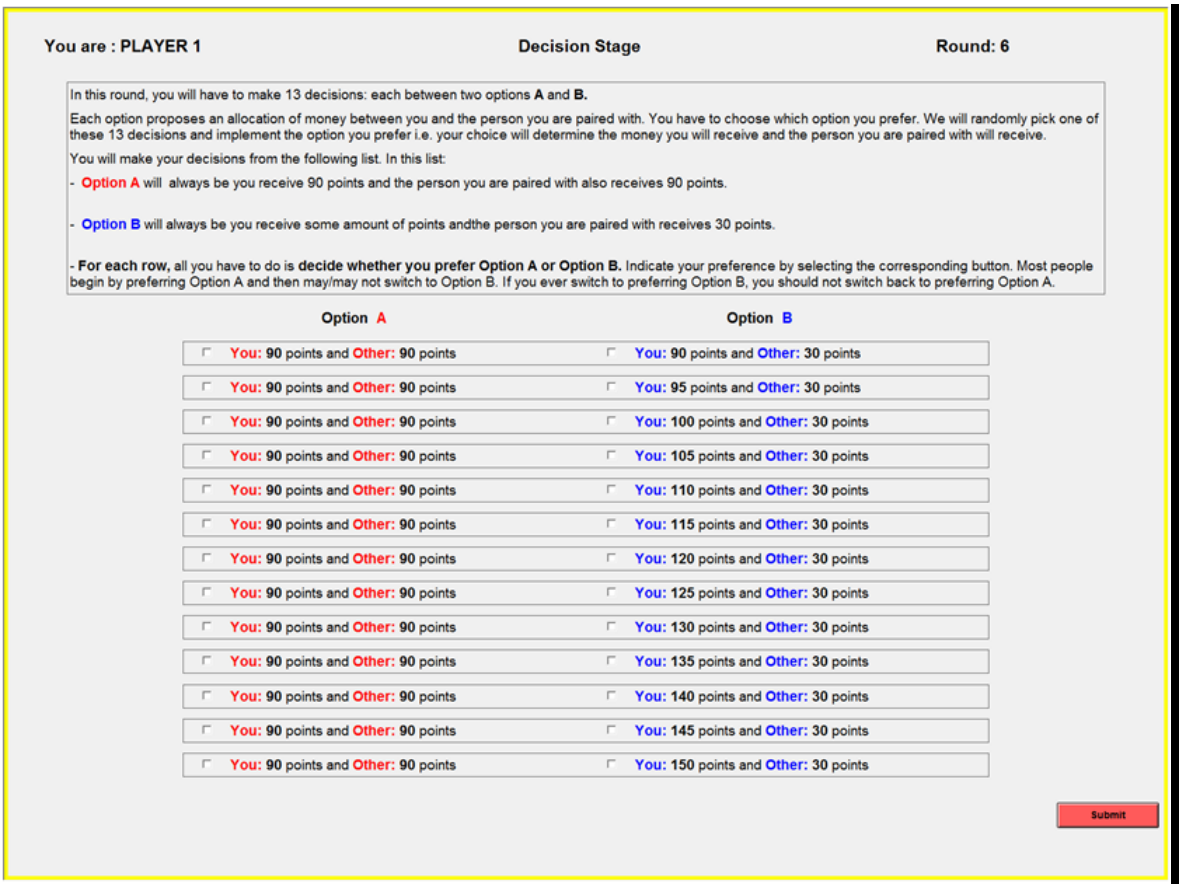



Figure B3. : Dictator Decisions

that workers encoding less than the median number of letters switch from preferring Option A to preferring Option B significantly earlier than workers encoding more than the median, indicating more selfish preferences ($p$-value=0.05). Table B1 shows that including a variable denoting the switch point has a positive and significant effect on the number of letters encoded for the Joint Project, confirming our conjecture that lower work done for the Joint Project correlates with lower baseline altruism. If it is the case that the distribution of work for the Joint Project simply reflects the distribution of subjects' altruistic preferences, it could well be that the difference in work observed *between* treatments is due to differences in the levels of altruism. Though our random assignment of subjects should eliminate systematic differences across treatments, we compare the aver-

Table B1—: Decision to Work

| Dependent variable: Number of letters encoded for Joint Projet | | | |
|---|---|---|---|
| | Full Sample (All sessions) | Sub-sample (Sessions 10-16) | Sub-sample (Sessions 10-16) |
| **Message Before** | -8.38** | -14.14** | -12.35** |
| | (4.20) | (6.17) | (6.02) |
| **Part 1 Performance** | 0.39 | 0.18 | 0.16 |
| | (0.26) | (0.40) | (0.41) |
| **Baseline Altruism** | | | 1.45** |
| | | | (0.70) |
| Constant | 24.51 | 54.13* | 42.45 |
| | (18.39) | (29.67) | (30.37) |
| Controls | Yes | Yes | Yes |
| No. of Obs. | 190 | 88 | 88 |
| R-Squared | 0.10 | 0.14 | 0.18 |

Numbers in parenthesis are robust standard errors. Controls include dummy if the participant's native language is English, dummies for year in school and a dummy for graduate student. Message Before is a treatment dummy variable (1=*MB*, 0=*MA*), Part 1 Performance is the subject's performance in the Part 1 calibration, and Baseline Altruism is measured from the subject's dictator game decision.
* $p$ <0.10, ** $p$ <0.05, *** $p$ <0.01.

age switch points in each treatment. We find no difference in subject's altruism between them ($p$-value=0.28). Hence, our results suggest that timing of communication predominantly affects behavior for individuals who are less altruistic.

One possible explanation of our results is that less altruistic individuals have a larger margin to deceive in persuading the manager to invest as compared to individuals who are more altruistic, and if the timing of communication affects costs of misinforming differently, one should expect it to affect the lower-tail more than the upper tail. Indeed in our data, subjects below the median work less in the *MB* treatment, but their messages are significantly *higher* (65.9 letters versus 50.9 letters, $p$-value=0.002). But for the above median workers, there is no difference in the messages sent (71.4 vs 71.1, $p$-value=0.95).

### *B4. Matrix Games*

Building on the results from Section IV.B, we report results from all five matrix games used. The games are depicted in Table B2. The Manager-Worker game was discussed in the text, and we include it here for comparison with the Prisoner's Dilemma and Stag Hunt games. In both Manager-Worker and Prisoner's Dilemma games, we also included a version of the game where the temptation payoff was reduced from 130 to 110.

Table B2—: Matrix Games

| | | **Receiver** | |
|---|---|---|---|
| | | *D* | *C* |
| **Sender** | *D* | 70, 70 | 130, 30 |
| | *C* | 30,80 | 90, 90 |

(a) Manager-Worker

| | | **Receiver** | |
|---|---|---|---|
| | | *D* | *C* |
| **Sender** | *D* | 70, 70 | 130, 30 |
| | *C* | 30, 130 | 90, 90 |

(b) Prisoner's Dilemma

| | | **Receiver** | |
|---|---|---|---|
| | | *D* | *C* |
| **Sender** | *D* | 70, 70 | 80, 30 |
| | *C* | 30,80 | 90, 90 |

(c) Stag Hunt

A motivation in using these other games is in knowing whether our results generalize to other environments involving strategic information transmission. We use a Prisoner's Dilemma and a Stag Hunt game for two reasons. First, in both the games, conditional on wanting to choose $D$, senders have an incentive to assure the receiver that he has chosen (will choose) $C$.[42] This is crucial to our environment as we want senders to have a monetary incentive to deceive their counterparts.

Second, in both games the efficient outcome occurs when both players choose $C$, similar to the manager-worker game. However, the games differ in the reason why

[42]In both games, regardless of his action, the sender prefers the receiver choose $C$. In the stag hunt game, the receiver does better by choosing $C$ *only* if the sender has also chosen $C$. In the prisoner's dilemma, if receivers are other-regarding, they may choose $C$ to an expectation of the sender choosing $C$. Hence, in both games, if the sender wants the receiver to choose $C$, he has to persuade the receiver that he has chosen $C$ with sufficiently high probability. Additionally, $D$ is the sender's dominant strategy in the prisoner's dilemma while in the stag hunt $D$ is the sender's risk-dominant strategy. This creates an incentive for the sender to choose $D$ in both games.

players may fail to achieve it. This allows us to investigate whether our results extend to situations strategically different from the manager-worker game. In the prisoner's dilemma, the sender has a dominant strategy to choose $D$ identical to the manager-worker game. In contrast, the receiver no longer has an incentive to coordinate as she now also has a dominant strategy of choosing $D$. In the stag hunt game, both players want to coordinate their actions on the payoff dominant outcome $(C,C)$, but $(D,D)$ is the risk-dominant outcome. Furthermore, note that the only difference between the games is in the payoff a player receives if he chooses $D$ and the other player chooses $C$ (temptation payoff).[43] Starting from the manager-worker game, we increase the receiver's temptation payoff to get the prisoner's dilemma, and decrease the sender's temptation to get the stag hunt.

Third, we use the Stag Hunt game to compare with existing theoretical predictions. Zultan (2013) and Schlag and Vida (2015) propose solution concepts that predict truthful communication equilibria in the *MB* but not *MA* treatment. We look to see how this interacts with other forces driving differences between *MB* and *MA*.

### *B5. Results*

Figure B4 indicates the fraction of senders who misinform in each of the five games across treatments. A considerable fraction of senders send false signals in the manager-worker and prisoner's dilemma games. In contrast, only a few do so in the stag hunt. This is expected as both players do better by coordinating their actions in the stag hunt. In the high stakes games, we find strong confirmation of our previous results; senders deceive significantly more in the *MB* than the *MA* treatment. However, the data fail to support these hypotheses when considering the low stake games. These results are formalized in a regression in Table B3.

In the stag-hunt game, we find no effect of timing on sender behavior. This observation is in contrast to the experimental findings of Charness (2000). Over ten rounds of a repeated stag hunt game, Charness (2000) finds senders who send a message before taking an action deceive only 5.8% of times as compared to 35% when senders send a message after taking an action. We find 18% and

[43]In the stag hunt, both players' temptation payoff is lower than the payoff from mutual cooperation. In the manager-worker, the receiver's temptation payoff is lower while the sender's is higher compared to mutual cooperation. In the prisoner's dilemma, both players' temptation payoff is higher than the payoff from mutual cooperation.

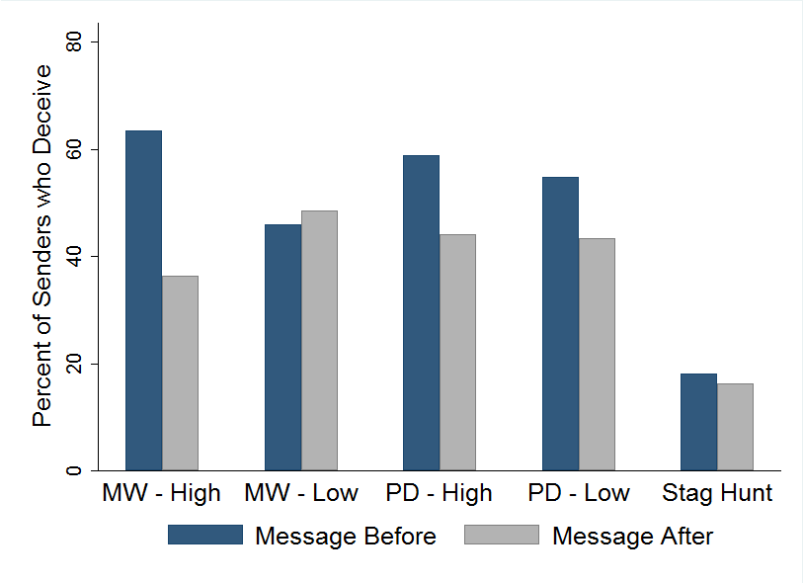



Figure B4. : Misinformation in the Matrix Games

*Note:* Percent of Senders who deceive i.e. senders who signal the cooperative action $C$ but choose the selfish action $D$. *MW:* Manager Worker; *PD:* Prisoner's Dilemma; *High* and *Low* refers to the value of the temptation payoff. *High Stake:* temptation payoff 130; Low Stake: temptation payoff 110.

16% of senders deceive in the *MB* and *MA* treatment respectively. One possible explanation could be that Charness (2000) employs a ten-round repeated game with feedback between rounds, while in our design subjects play the stag hunt only once. As the author mentions, cooperative play decreases significantly over time in the *MA* treatment, and comparing our results to the round one decisions in Charness (2000) would be a more apt comparison. Additionally, our results from the *MB* treatment can be compared to Clark, Kay and Sefton (2001), who find similar results from a stag hunt game with pre-play communication (albeit two-way communication). Our observations are comparable to their results.

These results show the difference between ex-ante and ex-post communication may extend to more general domains. Our results are robust to restricting the sample to exclude those who send no message, as well as when we compare the fraction of senders who send truthful messages.

Decomposing Misinformation in Matrix Games. — Analogous to the real-effort task, we can decompose the misinformation in the Choice Task into its two com-

Table B3—: Probit predicting whether sender misinforms in the Choice Task

| Dependent variable: Probability Sender Misinforms | | | | | |
|---|---|---|---|---|---|
| | Manager-Worker | | Prisoner's Dilemma | | Stag Hunt |
| | High | Low | High | Low | |
| | (1) | (2) | (3) | (4) | (5) |
| **Message Before** | 0.24*** | 0.04 | 0.16** | 0.13 | 0.05 |
| | (0.07) | (0.08) | (0.07) | (0.08) | (0.06) |
| **Round** | 0.02 | 0.05** | 0.06** | 0.03 | 0.04 |
| | (0.03) | (0.03) | (0.03) | (0.03) | (0.03) |
| **Worker** | -0.01 | 0.06 | 0.04 | 0.10 | 0.09 |
| | (0.10) | (0.10) | (0.10) | (0.11) | (0.08) |
| No. of Obs. | 142 | 142 | 142 | 142 | 142 |

*Note:* * $p$ <0.10, ** $p$ <0.05, *** $p$ <0.01. Numbers in parenthesis are robust standard errors clustered at subject-level. Controls include participant's age, gender dummy and race dummies. Message Before is a treatment dummy variable (1=*MB*, 0=*MA*), Worker is a Part 2 role dummy variable (1=worker, 0=manager), and Round is the order in which the game appeared. Marginal effects reported.

ponents - signal and action. Due to the binary nature of our decision variables, we compare the frequency of cooperative signals and actions across treatments. Table B4 displays the overall fraction of senders who signal the cooperative action $C$ and the fraction of senders who choose $C$. Recall, in the real-effort task, messages sent before the action is taken promised a higher level of cooperation. Table B4 Columns 3-5 indicates that same pattern. A vast majority of senders choose to signal $C$ in both treatments. Additionally the percent of $C$ signals is always higher in the *MB* treatment for all games. However, Table B4 shows these differences are significant only for the high-stake manager-worker and stag-hunt games. The $p$-value for the high-stake prisoner's dilemma is borderline significant at 0.11.[44] Turning to the fraction of times senders actually took the cooperative action $C$, Columns 6-8 indicate that, except for the stag hunt and manager-worker low stake, senders are more likely to choose the cooperative action. Moreover, the magnitudes on the coefficients are very small. Instead, senders are more cooperative in the *MB* treatment in the stag hunt game. This latter result provides evidence for Farrell's argument and is consistent with Charness [2000] results. It

[44]The $p$-values for prisoner's dilemma high and low stake are 0.11 and 0.19 respectively. Since, our primary design was the real-effort task, our power calculation was based on it, we are probably underpowered to conclusively identify a treatment effect with binary variables. If we pool the data for the two stakes in the prisoner's dilemma, the treatment effect is significant at 10%, which provides some supporting evidence for the sample being underpowered.

also proves that participants are responding to incentives and the structure of the strategic interaction.

Table B4—: Signal and Actions

| | | **Percentage of Senders sending Signal *C*** | | | **Percentage of Senders choosing Action *C*** | | |
|---|---|---|---|---|---|---|---|
| | | Message Before | Message After | $p$-value | Message Before | Message After | $p$-value |
| Manager-Worker | High | 91.9 | 75.0 | 0.006 | 29.7 | 44.1 | 0.07 |
| | Low | 82.4 | 80.9 | 0.81 | 36.5 | 32.3 | 0.61 |
| Prisoner's Dilemma | High | 79.7 | 73.5 | 0.38 | 24.3 | 30.9 | 0.38 |
| | Low | 82.4 | 72.1 | 0.14 | 29.7 | 32.4 | 0.74 |
| Stag Hunt | | 89.2 | 76.5 | 0.04 | 75.7 | 60.3 | 0.05 |

# Appendix C: Instructions

# Instructions (NOT FOR PUBLICATION)

## Welcome!

Thank you for participating in this study. This is a study of individual behavior and decision making. At this time, please turn off and put away any electronic devices/phones you may have brought with you. There may be moments where you will have to sit and wait while others in the room make their decisions, and we ask you to be patient.

This experiment has 3 parts. In each part you have the opportunity to earn points. The points will be converted to dollars for your final payment. The conversion rate is $1 = 10 points. In addition, you can earn points in a bonus stage. One part will be randomly chosen to determine your payment at the end of the session.

## Part 1

In Part 1, everyone will have a chance to earn money by working on a task. The task is the same for everyone. We will call it the *Encoding Task*. The task consists of converting letters into numbers. Your screen displays a table with two rows. The first row contains all of the letters in the alphabet and the second row provides a number that goes with that letter. During the task, you will be given a letter and you must enter the corresponding number in the box on your screen. You must validate your answer by pressing the 'Submit' button. The computer only accepts correct entries, so if you answer incorrectly, a prompt will ask you to correct it. Once you submit the correct entry for that letter, the table will reset and you will be presented with another letter to encode and so on. You will have three minutes to convert as many letters as you can. A counter on the screen will keep track of the number of letters you encode.

- At the end of the 3 minutes, you will be informed of the number of letters you have encoded, broken down per minute.

**Earnings for Part 1:**

If Part 1 is chosen for payment, you will be paid based on how many letters you encode. There is a total amount of 100 points. The total number of letters you encode determines the *share of these 100 points* you will receive, in the following way:

Imagine there is a bag with 23 blue balls in it. For every letter you encode, we put a red ball into the bag. This means, if you encode a total of 'x' number of letters in the 3 minutes, the bag will contain 23 blue balls + 'x' red balls at the end of the 3 minutes.
The share of the 100 points you earn will be the percentage of red balls in the bag, given by:

$$=\frac{x}{x+23} * 100$$

If you encode 0 letters, you receive $\frac{0}{0+23} * 100$**=0 points**

If you encode 1 letter, you receive $\frac{1}{1+23} * 100 = \frac{100}{24}$ **= 4.17 points**

If you encode 2 letters, you receive $\frac{2}{2+23} * 100 = \frac{200}{25} = 8$ **points**

….and so on…

Notice that the more letters you encode, you ***always*** earn more as the share of red balls in the bag increases. However, every *additional* letter you encode gives you a lower share per letter than the previous.

The exact payment schedule is listed in the table below: (Please take a moment to go over this, and ask us any questions you may have now).

Before we start, you will be given a chance to practice this task for a minute to familiarize yourself with the task. The number of letters converted during this practice time will not affect your earnings.

# Part 2

In this part, you will be randomly assigned either the role of **A** or the role of **B**. The decisions you make may affect your earnings and the earnings of others.

*Overview*

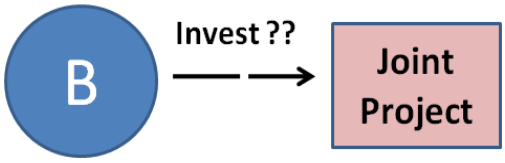
Each Player A will interact with another randomly chosen Player B in this room. *The amount of points you earn depends on the decisions made in your pair.* Your interaction is completely anonymous, so participants will only be referred to as A and B for the duration of the experiment.

Here is what you will have to do:

***Player B Decision and Payoff:***

Player B decides whether to invest or not in a project called "Joint Project". Investing is profitable *only* if the Joint Project is "successful".

- If B **invests**, and the Joint Project is **successful**, she receives **130 points**.
- If B **invests**, and the Joint Project is **unsuccessful**, she receives **10 points**.
- If B does **not invest** she receives **70 points**.




*The success of the Joint Project, however is not under Player B's control. It depends on Player A and how much Player A works for the Joint Project.* The more Player A works for the Joint Project, the higher the chance of it being successful (more on this below).

***Player A Decision and Payoff:***

Player A has two projects he can work on – the "Joint Project" and the "Personal Project". He has 4 minutes to work and he can split his time anyway he likes between the two Projects. Both Projects entail working on a series of encoding tasks as before. The number of letters encoded for each project is the measure of work done for that project. Player A's earnings is a sum of his earnings from each project.

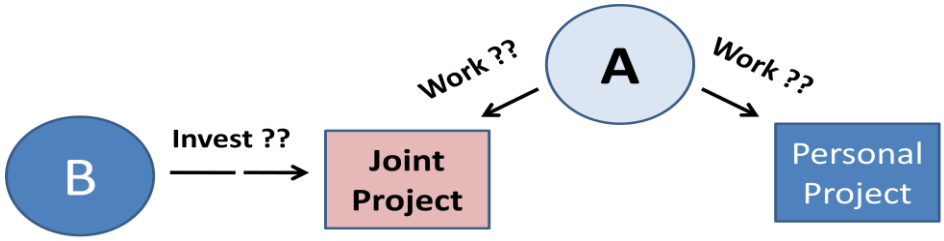
- Joint Project Earnings: Player A's earnings from the Joint Project depends *only* on Player B's investment decision.
  - **If Player B invests** in the Joint Project, Player A earns **120 points** from the Joint Project.

- **If Player B does not invest** in the Joint Project, Player A receives **0 points** from the Joint Project.
- These earnings do NOT depend on whether the Joint Project is successful.

➢ Personal Project Earnings: Player A's earnings from the Personal Project depends on how much he works (number of letters encoded) for the Personal Project. The payment for the Personal Project is exactly the same payment schedule as in Part 1.

- There is a total of 100 points. The total number of letters Player A encodes for the Personal Project determines the ***share of these 100 points*** **he receives**. There is a bag labeled "Personal Bag" which contains 23 blue balls. For every letter Player A encodes for the Personal Project, we add a red ball to the Personal Bag. After the 4 minutes are over, Player A receives a share from these 100 points. The percentage of red balls in the Personal Bag determines the share he receives.

Player A's Earnings = Earnings from Joint Project **+** Earnings from Personal Project.




***When is the Joint Project Successful?***

The amount of work done by Player A for the Joint Project (number of letters encoded for the Joint Project) determines the success of the Joint Project in the following way:

There is another bag labeled 'Joint Bag' which also contains 23 blue balls. For every letter Player A encodes for the Joint Project, we put a red ball in the 'Joint Bag'. After the end of the 4 minutes work time, we randomly pick a ball from this bag. If the ball drawn is red, the Joint Project is successful. If it is blue, the Joint Project fails.
This means that the chance of Joint Project being successful increases with the number of letters encoded by Player A for the Joint Project, as it increases the number of red balls in the 'Joint Bag' and makes it more likely that a red ball is selected. Conversely, the less Player A encodes the higher the chance that the Joint Project fails.

To represent this mathematically, if Player A encodes '*j*' number of letters for the Joint Project, the exact chance of success for the Joint Project is given by:

$$= \frac{j}{j+23} * 100$$

A sheet is provided to you that list the chance of success of the Joint Project for *each* possible number of letters encoded by Player A for the Joint Project. It also lists the earnings Player A would receive for the corresponding number of letters encoded in the Personal Project (if you didn't get the sheet, please raise your hand).

***Information:***

Player A will have 4 minutes to work on his Projects. Then, after the work-stage is over, Player B will make her investment decision.Note that Player B does NOT learn whether the Project is successful until AFTER she makes the investment decision. Player B NEVER finds out how many letters were encoded by Player A for the Joint Project; she only comes to know if it was successful or not.

***Message:***

After the work stage (but before Player B makes her investment decision), Player A has an option to send a message to Player B. In this message, Player A can state how many letters he has encoded for the Joint Project. Player B receives this message, after which Player B makes her investment choice. *The message is the only information Player B receives before she makes her investment decision.*

***The sequence of the decisions is as follows:***

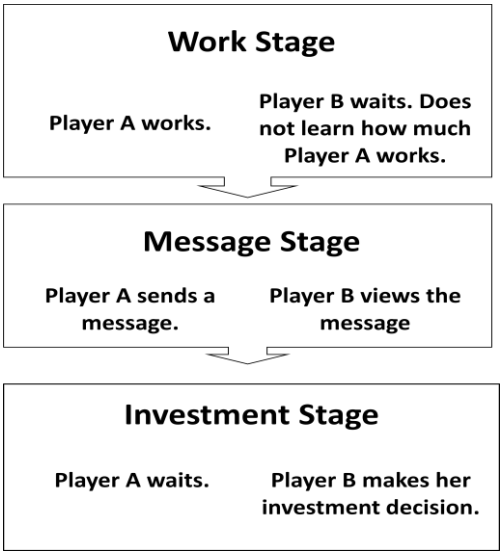
1. Roles determined randomly in a pair. Participants informed of their roles and the ID of their co-participant.

2. **Work Stage:** Player A will have 4 minutes to work. For every letter encoded in the work-stage, Player A can decide which Project he would want the work to go towards. Player A will select which Project he wants to start with. At any point during the work-stage A will be allowed to switch back and forth as often he likes between the two Projects. There is no limit on the amount of times he is allowed to switch between these two Projects. Two counters displayed on the screen will keep track of the number of letters encoded for *each* Project. Between these, there are buttons that will allow Player A to switch between working for the two projects. If you are unsure about which Project your current work is going towards, you can find this information directly between the counters.

1. **Message Stage:** Player A will be given the opportunity to send a message to Player B, where A can fill in a statement regarding how many letters he has encoded for the Joint Project as well as a suggestion to Player B on her action. The statement reads "Hi, I have encoded ____ letters for the Joint Project. You should/should not invest." Player A can

fill in any non-negative number in the blank. If Player A chooses not to send any message, Player B will receive the message "A has chosen not to send any message". The message from A will be communicated to B as soon as he sends it.

2. **Investment Stage:** After Player B receives a message (if Player A sends one), Player B makes her investment choice. B will be able to see the message A sent her when she makes her investment decision.




We will provide the following information about your co-participant's decision and the outcome **at the end of the session**.

- Player A will be informed whether Player B invested or not and if the Joint Project was successful.
- If Player B **invested** in the Joint Project, she will only *be informed whether the Project was a success or a failure*. *Player B will not be informed of the number of letters encoded by Player A for the Joint Project.* If Player B **does not invest** in the Joint Project, she receives no information about the outcome of the Joint Project.

The last page is a summary broken down by role. There are a few questions about the procedure after that to test your understanding of the instructions. Please review that and raise your hand if you have a question. We'll take a few minutes to answer all questions, and then we'll begin.

To recap, broken down by roles:

***Player A***

- Decides to split 4 minutes worth of work between Personal Project and Joint Project.
- Can "send a message" to Player B after the Work Stage about how much he has worked for the Joint Project.
- Receives $\frac{x}{x+23} * 100$ points for 'x' letters encoded for the Personal Project.
- Receives **120 points** from the Joint Project, *only* if Player B invests in the Joint Project, otherwise **0 points.**

***Player B***

- Decides whether or not to invest in the Joint Project.
- "Receives a message" from Player A before she makes her investment decision, but never learns the number of letters encoded for the Joint Project
- If she invests and Joint Project is
  - successful receives **130 points**
  - failure receives **10 points**
- If she does not invest receives **70 points**.

1. Assume that you are **Player B** and you invest in the Joint Project. If the Joint Project is successful, you will receive _____ points, and if the Joint Project fails, you will receive _____ points.
2. Assume that you are **Player B** and you do not invest in the Joint Project. If the Joint Project is successful, you will receive _____ points, and if the Joint Project fails, you will receive _____ points.

3. The success of the Joint Project depends on which of the following?
   a. The number of letters Player A encodes for the Personal Project.
   b. The number of letters Player A encodes for the Joint Project.
   c. Player B's investment decision.

4. If you are **Player B,** you
   a. NEVER get to see how many letters Player A encoded for the Joint Project. (TRUE/FALSE)

b. will get to see the success/failure of the Joint Project BEFORE you decide whether or not to invest in the Joint Project. _____ (TRUE/FALSE)

5. will get to see the message from Player A (if Player A chooses to send a message), BEFORE you decide whether or not to invest in the Joint Project. _____ (TRUE/FALSE)

6. Does Player A's payoff from the Joint Project depend upon

   a. The success of the Joint Project? (YES/NO)
   b. Player B's investment decision? (YES/NO)
   c. How many letters Player A encodes for the Joint Project? (YES/NO)

7. Does Player A's payoff from the Personal Project depend upon

   a. The success of the Joint Project? (YES/NO)
   b. Player B's investment decision? (YES/NO)
   c. How many letters Player A encodes for the Personal Project? (YES/NO)

# Part 3

This is the final part of the experiment. In this part, you will be asked to make decisions in 5 different rounds.

**Roles:**

- At the beginning of Round 1, half of the people in this room will be assigned the role of **Player 1** and the other half will be assigned the role of **Player 2**. Your role will remain constant over all 5 rounds.
- In each round, a Player 1 will be matched with a Player 2, and the two of you will form a pair. You will be matched with a DIFFERENT person in the room in each round, *so you will never interact with the same person more than once*.
- In each round, Player 1 will make his choice **first**, and Player 2 will make her choice **second**.
- Player 2 will NOT learn Player 1's decision before she makes her own choice.

**Decisions and Payoffs:**

In each round, you will decide whether you want to choose action '**X**' or '**Y**.' Your payoffs will depend on what *you* choose and on what the *other perso*n in your pair chooses. The possible payoffs depending on the choice both of you make will be represented in a table, like this –

| | | PLAYER 2 | |
|---|---|---|---|
| | | **X** | **Y** |
| PLAYER 1 | **X** | 70,70 | 80,30 |
| | **Y** | 30,80 | 90,90 |

In each cell, Player 1's payoff is listed first and Player 2's payoff is listed second. The actual payoffs that are realized will be determined by Player 1's choice and Player 2's choice together. You can read the payoffs from the table in this way:

- If Player 1 choses **X** and Player 2 chooses **X**, each player receives 70 points.
- If Player 1 chooses **X** and Player 2 chooses **Y**, Player 1 receives 80 points and Player 2 receives 30 points.
- If Player 1 chooses **Y** and Player 2 chooses **X**, Player 1 receives 30 points and Player 2 receives 80 points.
- If Player 1 chooses **Y** and Player 2 chooses **Y**, each player receives 90 points.

To make the tables easier to read, the computer will always highlight *your* payoffs in blue while you're making your decisions.

In each round, the payoffs from your choices (or numbers in the cells) will be DIFFERENT. So please carefully read the payoffs from the table in each round before making a choice. The actual tables you will see are all listed on the last page.

**Message:**

Player 1 will have the opportunity to send a message to Player 2 after Player 1 makes his choice, but before Player 2 takes her action. In the message, Player 1 can indicate which choice he made, as well as a suggestion to Player 2 on her action. Player 1 can send one of three messages: "I have chosen X/Y. You should choose X", "I have chosen X/Y. You should choose Y," or "No message." The selected message will be delivered from Player 1 to Player 2, and then Player 2 will make her choice: X or Y. Player 2 will NOT see whether Player 1 chose 'X' or 'Y' before she makes her choice. She will only see the message.

**The sequence of Part 3 is as follows:**

1. **Player 1 Decision:** Within a pair, Player 1 will make his decision and choose X or Y. This choice will not be revealed to Player 2.
2. **Message stage:** Player 1 will be given the opportunity to send a message to Player 2, where Player 1 can make a statement about the choice he has made and a suggestion to Player 2 on what she should choose. Once Player 1 sends this message, it is delivered to Player 2.
3. **Player 2 Decision:** Player 2 will make her decision and choose X or Y.
4. **New round begins:** You will be randomly matched with another participant, and will go through this sequence again. You will play a total of 5 rounds. You will stay in the same role as Player 1 or Player 2 for all 5 rounds.

At the end of the 5 rounds, you will be asked to fill out a short questionnaire.

We will provide the following information about decisions made at **the end of the session**:your partner's choice in each round and your payoff in each round**.** You will not learn your payoffs from any round until the very end of the session.

**Part 3 Payoffs:**

If Part 3 is chosen for payment, the computer will randomly select 1 of the 5 rounds and the points you earn from that round will determine your payment. Since, there is no way to know which round the computer will select, you should make your decision in each round as if it will determine your payment.

**Final Payment:**

One of the three parts will be randomly picked for payment. You will receive $5 as a show-up fee in addition to your earnings from that part.

The following are the 5 games you will see, one in each round. The computer randomly selects one game for every round, so the sequence might not be the same as listed here. In each table at least one payoff changes which alters the game. So pay attention and make your decision wisely.

**Game 1**

| | | PLAYER 2 | |
|---|---|---|---|
| PLAYER 1 | | **X** | **Y** |
| | **X** | 70,70 | 110,30 |
| | **Y** | 30,110 | 90,90 |

**Game 2**

| | | PLAYER 2 | |
|---|---|---|---|
| PLAYER 1 | | **X** | **Y** |
| | **X** | 70,70 | 110,30 |
| | **Y** | 30,80 | 90,90 |

**Game 3**

| | | PLAYER 2 | |
|---|---|---|---|
| PLAYER 1 | | **X** | **Y** |
| | **X** | 70,70 | 80,30 |
| | **Y** | 30,80 | 90,90 |

**Game 4**

| | | PLAYER 2 | |
|---|---|---|---|
| | | **X** | **Y** |
| PLAYER 1 | **X** | 70,70 | 130,30 |
| | **Y** | 30,80 | 90,90 |

**Game 5**

| | | PLAYER 2 | |
|---|---|---|---|
| | | **X** | **Y** |
| PLAYER 1 | **X** | 70,70 | 130,30 |
| | **Y** | 30,130 | 90,90 |